

SEQUENCE LISTING

```

#          SEQUENCE LISTING          - - - - (1) GENERAL INFORMATION:
- -      (iii) NUMBER OF SEQUENCES: 26
- - - - (2) INFORMATION FOR SEQ ID NO:1:
- -      (i) SEQUENCE CHARACTERISTICS:
- -          (A) LENGTH: 1057 base - #pairs
- -          (B) TYPE: nucleic acid              (C) STRANDEDNESS: double
- -          (D) TOPOLOGY: linear                - -      (ix) FEATURE:
- -          (A) NAME/KEY: CDS                    (B) LOCATION: 124..893
- -      (ix) FEATURE:                          (A) NAME/KEY: misc.sub.-- - #feature
- -          (B) LOCATION: 1..1057
- -          (D) OTHER INFORMATION: - #/note= "product = Arabidopsis
- -              thaliana - #AP1."
- -      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
- -      CTTTCCAATT GGTCATACC AAAGTCTGAG CTCTTCTTTA TATCTCTCTT GT -
#AGTTTCTT      60
- -      ATTGGGGGTC TTTGTTTTGT TTGGTTCTTT TAGAGTAAGA AGTTTCTTAA AA -
#AAGGATCA      120
- -      AAA ATG GCA AGG GGT AGG GTT CAA TTG AAG AG - #G ATA GAG AAC AAG
ATC      168
      Met Gly Arg Gly Arg Val Gln Leu - #Lys Arg Ile Glu Asn Lys Ile
      1          - # 5          - # 10          - # 15
- -      AAT AGA CAA GTG ACA TTC TCG AAA AGA AGA GC - #T GGT CTT TTG AAG AAA
      216
      Asn Arg Gln Val Thr Phe Ser Lys Arg Arg Al - #a Gly Leu Leu Lys Lys
      20 - #          25 - #          30
- -      GCT CAT GAG ATC TCT GTT CTC TGT GAT GCT GA - #A GTT GCT CTT GTT GTC
      264
      Ala His Glu Ile Ser Val Leu Cys Asp Ala Gl - #u Val Ala Leu Val Val
      35 - #          40 - #          45
- -      TTC TCC CAT AAG GGG AAA CTC TTC GAA TAC TC - #C ACT GAT TCT TGT ATG
      312
      Phe Ser His Lys Gly Lys Leu Phe Glu Tyr Se - #r Thr Asp Ser Cys Met
      50          - #          55          - #          60
- -      GAG AAG ATA CTT GAA CGC TAT GAG AGG TAC TC - #T TAC GCC GAA AGA CAG
      360
      Glu Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Se - #r Tyr Ala Glu Arg Gln
      65          - #          70          - #          75
- -      CTT ATT GCA CCT GAG TCC GAC GTC AAT ACA AA - #C TGG TCG ATG GAG TAT
      408
      Leu Ile Ala Pro Glu Ser Asp Val Asn Thr As - #n Trp Ser Met Glu Tyr
      80          - # 85          - # 90          - # 95
- -      AAC AGG CTT AAG GCT AAG ATT GAG CTT TTG GA - #G AGA AAC CAG AGG CAT
      456
      Asn Arg Leu Lys Ala Lys Ile Glu Leu Leu Gl - #u Arg Asn Gln Arg His
      100 - #          105 - #          110
- -      TAT CTT GGG GAA GAC TTG CAA GCA ATG AGC CC - #T AAA GAG CTT CAG AAT
      504
      Tyr Leu Gly Glu Asp Leu Gln Ala Met Ser Pr - #o Lys Glu Leu Gln Asn
      115          - #          120          - #          125

```

- - CTG GAG CAG CAG CTT GAC ACT GCT CTT AAG CA - #C ATC CGC ACT AGA AAA
552
Leu Glu Gln Gln Leu Asp Thr Ala Leu Lys Hi - #s Ile Arg Thr Arg Lys
130 - # 135 - # 140
5 - - AAC CAA CTT ATG TAC GAG TCC ATC AAT GAG CT - #C CAA AAA AAG GAG AAG
600
Asn Gln Leu Met Tyr Glu Ser Ile Asn Glu Le - #u Gln Lys Lys Glu Lys
145 - # 150 - # 155
10 - - GCC ATA CAG GAG CAA AAC AGC ATG CTT TCT AA - #A CAG ATC AAG GAG AGG
648
Ala Ile Gln Glu Gln Asn Ser Met Leu Ser Ly - #s Gln Ile Lys Glu Arg
160 1 - #65 1 - #70 1 -
#75
- - GAA AAA ATT CTT AGG GCT CAA CAG GAG CAG TG - #G GAT CAG CAG AAC
15 CAA 696
Glu Lys Ile Leu Arg Ala Gln Gln Glu Gln Tr - #p Asp Gln Gln Asn Gln
180 - # 185 - # 190
- - GGC CAC AAT ATG CCT CCC CCT CTG CCA CCG CA - #G CAG CAC CAA ATC CAG
744
20 Gly His Asn Met Pro Pro Pro Leu Pro Pro Gl - #n Gln His Gln Ile Gln
195 - # 200 - # 205
- - CAT CCT TAC ATG CTC TCT CAT CAG CCA TCT CC - #T TTT CTC AAC ATG GGT
792
His Pro Tyr Met Leu Ser His Gln Pro Ser Pr - #o Phe Leu Asn Met Gly
25 210 - # 215 - # 220
- - GGT CTG TAT CAA GAA GAT GAT CCA ATG GCA AT - #G AGG AGG AAT GAT CTC
840
Gly Leu Tyr Gln Glu Asp Asp Pro Met Ala Me - #t Arg Arg Asn Asp Leu
225 - # 230 - # 235
30 - - GAA CTG ACT CTT GAA CCC GTT TAC AAC TGC AA - #C CTT GGC TGC TTC GCC
888
Glu Leu Thr Leu Glu Pro Val Tyr Asn Cys As - #n Leu Gly Cys Phe Ala
240 2 - #45 2 - #50 2 -
#55
35 - - GCA TG AAGCATTTCC ATATATATAT TTGTAATCGT CAACAATAAA AAC - #AGTTTGC
943 Ala
- - CACATACATA TAAATAGTGG CTAGGCTCTT TTCATCCAAT TAATATATTT TG -
#GCAAATGT 1003
- - TCGATGTTCT TATATCATCA TATATAAATT AGCAGGCTCC TTTCTTTTTT TG - #TA
40 1057 - - - (2) INFORMATION FOR SEQ ID NO:2:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
45 - - Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
20 - # 25 - # 30
- - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
50 35 - # 40 - # 45
- - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Th - #r Asp Ser Cys Met Glu

```

50          - #      55          - #      60
- - Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
65          - # 70          - # 75          - # 80
- - Ile Ala Pro Glu Ser Asp Val Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
5          85 - #          90 - #          95
- - Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
          100 - #          105 - #          110
- - Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
          115 - #          120 - #          125
10 - - Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Thr Arg Lys Asn
          130 - #          135 - #          140
- - Gln Leu Met Tyr Glu Ser Ile Asn Glu Leu Gl - #n Lys Lys Glu Lys Ala
145          1 - #50          1 - #55          1 -
#60
15 - - Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg
Glu          165 - #          170 - #          175
- - Lys Ile Leu Arg Ala Gln Gln Glu Gln Trp As - #p Gln Gln Asn Gln Gly
          180 - #          185 - #          190
- - His Asn Met Pro Pro Pro Leu Pro Pro Gln Gl - #n His Gln Ile Gln His
20          195 - #          200 - #          205
- - Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
          210 - #          215 - #          220
- - Leu Tyr Gln Glu Asp Asp Pro Met Ala Met Ar - #g Arg Asn Asp Leu Glu
225          2 - #30          2 - #35          2 -
25 #40
- - Leu Thr Leu Glu Pro Val Tyr Asn Cys Asn Le - #u Gly Cys Phe Ala
Ala          245 - #          250 - #          255
- - - (2) INFORMATION FOR SEQ ID NO:3:
- - (i) SEQUENCE CHARACTERISTICS:
- - (A) LENGTH: 794 base - #pairs
- - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
- - (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
- - (ix) FEATURE: (A) NAME/KEY: CDS
- - (B) LOCATION: 36..794 - - (ix) FEATURE:
35 (A) NAME/KEY: misc.sub.-- - #feature
- - (B) LOCATION: 1..794
- - (D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
AP1." - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
40 - - TCTTAGAGGA AATAGTTCCT TTAAAGGGA TAAAA ATG GGA AGG - #GGT AGG GTT
53
          - #          - # Met Gly Arg Gly Arg Val
          - #          - # 1          - # 5
- - CAG TTG AAG AGG ATA GAA AAC AAG ATC AAT AG - #A CAA GTG ACA TTC TCG
101
45 Gln Leu Lys Arg Ile Glu Asn Lys Ile Asn Ar - #g Gln Val Thr Phe Ser
          10 - #          15 - #          20
- - AAA AGA AGA GCT GGT CTT ATG AAG AAA GCT CA - #T GAG ATC TCT GTT CTG
149
Lys Arg Arg Ala Gly Leu Met Lys Lys Ala Hi - #s Glu Ile Ser Val Leu
50          25 - #          30 - #          35
- - TGT GAT GCT GAA GTT GCG CTT GTT GTC TTC TC - #C CAT AAG GGG AAA CTC

```

197
 Cys Asp Ala Glu Val Ala Leu Val Val Phe Se - #r His Lys Gly Lys Leu
 40 - # 45 - # 50
 - - TTT GAA TAC TCC ACT GAT TCT TGT ATG GAG AA - #G ATA CTT GAA CGC TAT
 245
 Phe Glu Tyr Ser Thr Asp Ser Cys Met Glu Ly - #s Ile Leu Glu Arg Tyr
 55 - # 60 - # 65 - # 70
 - - GAG AGA TAC TCT TAC GCC GAG AGA CAG CTT AT - #A GCA CCT GAG TCC GAC
 293
 Glu Arg Tyr Ser Tyr Ala Glu Arg Gln Leu Il - #e Ala Pro Glu Ser Asp
 75 - # 80 - # 85
 - - TCC AAT ACG AAC TGG TCG ATG GAG TAT AAT AG - #G CTT AAG GCT AAG ATT
 341
 Ser Asn Thr Asn Trp Ser Met Glu Tyr Asn Ar - #g Leu Lys Ala Lys Ile
 90 - # 95 - # 100
 - - GAG CTT TTG GAG AGA AAC CAG AGG CAC TAT CT - #T GGG GAA GAC TTG CAA
 389
 Glu Leu Leu Glu Arg Asn Gln Arg His Tyr Le - #u Gly Glu Asp Leu Gln
 105 - # 110 - # 115
 - - GCA ATG AGC CCT AAG GAA CTC CAG AAT CTA GA - #G CAA CAG CTT GAT ACT
 437
 Ala Met Ser Pro Lys Glu Leu Gln Asn Leu Gl - #u Gln Gln Leu Asp Thr
 120 - # 125 - # 130
 - - GCT CTT AAG CAC ATC CGC TCT AGA AAA AAC CA - #A CTT ATG TAC GAC TCC
 485
 Ala Leu Lys His Ile Arg Ser Arg Lys Asn Gl - #n Leu Met Tyr Asp Ser
 135 1 - #40 1 - #45 1 -
 #50
 - - ATC AAT GAG CTC CAA AGA AAG GAG AAA GCC AT - #A CAG GAA CAA AAC
 AGC 533
 Ile Asn Glu Leu Gln Arg Lys Glu Lys Ala Il - #e Gln Glu Gln Asn Ser
 155 - # 160 - # 165
 - - ATG CTT TCC AAG CAG ATT AAG GAG AGG GAA AA - #C GTT CTT AGG GCG CAA
 581
 Met Leu Ser Lys Gln Ile Lys Glu Arg Glu As - #n Val Leu Arg Ala Gln
 170 - # 175 - # 180
 - - CAA GAG CAA TGG GAC GAG CAG AAC CAT GGC CA - #T AAT ATG CCT CCG CCT
 629
 Gln Glu Gln Trp Asp Glu Gln Asn His Gly Hi - #s Asn Met Pro Pro Pro
 185 - # 190 - # 195
 - - CCA CCC CCG CAG CAG CAT CAA ATC CAG CAT CC - #T TAC ATG CTC TCT CAT
 677
 Pro Pro Pro Gln Gln His Gln Ile Gln His Pr - #o Tyr Met Leu Ser His
 200 - # 205 - # 210
 - - CAG CCA TCT CCT TTT CTC AAC ATG GGG GGG CT - #G TAT CAA GAA GAA GAT
 725
 Gln Pro Ser Pro Phe Leu Asn Met Gly Gly Le - #u Tyr Gln Glu Glu Asp
 215 2 - #20 2 - #25 2 -
 #30
 - - CAA ATG GCA ATG AGG AGG AAC GAT CTC GAT CT - #G TCT CTT GAA CCC
 GGT 773

Gln Met Ala Met Arg Arg Asn Asp Leu Asp Le - #u Ser Leu Glu Pro Gly
 235 - # 240 - # 245
 - - TAT AAC TGC AAT CTC GGC TGC - # - #
 794 Tyr Asn Cys Asn Leu Gly Cys 250
 5 - - - (2) INFORMATION FOR SEQ ID NO:4:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 253 amino - #acids (B) TYPE: amino acid
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
 10 - - Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Met Lys Lys Ala
 20 - # 25 - # 30
 15 - - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Th - #r Asp Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
 65 - # 70 - # 75 - # 80
 20 - - Ile Ala Pro Glu Ser Asp Ser Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
 85 - # 90 - # 95
 - - Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
 100 - # 105 - # 110
 - - Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
 115 - # 120 - # 125
 25 - - Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Ser Arg Lys Asn
 130 - # 135 - # 140
 - - Gln Leu Met Tyr Asp Ser Ile Asn Glu Leu Gl - #n Arg Lys Glu Lys Ala
 145 1 - #50 1 - #55 1 -
 #60
 - - Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg
 Glu
 35 - - Asn Val Leu Arg Ala Gln Gln Glu Gln Trp As - #p Glu Gln Asn His Gly
 180 - # 185 - # 190
 - - His Asn Met Pro Pro Pro Pro Pro Gln Gl - #n His Gln Ile Gln His
 195 - # 200 - # 205
 40 - - Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
 210 - # 215 - # 220
 - - Leu Tyr Gln Glu Glu Asp Gln Met Ala Met Ar - #g Arg Asn Asp Leu Asp
 225 2 - #30 2 - #35 2 -
 #40 - - Leu Ser Leu Glu Pro Gly Tyr Asn Cys Asn Le - #u Gly Cys
 245 - # 250
 45 - - - (2) INFORMATION FOR SEQ ID NO:5:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 768 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 50 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 1..766 - - (ix) FEATURE:

(A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..768
 (D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
 var. botr - #ytis AP1."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
 - - ATG GGA AGG GGT AGG GTT CAG TTG AAG AGG AT - #A GAA AAC AAG ATC AAT
 48
 Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - AGA CAA GTG ACA TTC TCG AAA AGA AGA GCT GG - #T CTT ATG AAG AAA GCT
 96
 Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Met Lys Lys Ala
 20 - # 25 - # 30
 - - CAT GAG ATC TCT GTT CTG TGT GAT GCT GAA GT - #T GCG CTT GTT GTC TTC
 144
 His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
 35 - # 40 - # 45
 - - TCC CAT AAG GGG AAA CTC TTT GAA TAC CCC AC - #T GAT TCT TGT ATG GAG
 192
 Ser His Lys Gly Lys Leu Phe Glu Tyr Pro Th - #r Asp Ser Cys Met Glu
 50 - # 55 - # 60
 - - GAG ATA CTT GAA CGC TAT GAG AGA TAC TCT TA - #C GCC GAG AGA CAG CTT
 240
 Glu Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
 65 - # 70 - # 75 - # 80
 - - ATA GCA CCT GAG TCC GAC TCC AAT ACG AAC TG - #G TCG ATG GAG TAT AAT
 288
 Ile Ala Pro Glu Ser Asp Ser Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
 85 - # 90 - # 95
 - - AGG CTT AAG GCT AAG ATT GAG CTT TTG GAG AG - #A AAC CAG AGG CAC TAT
 336
 Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
 100 - # 105 - # 110
 - - CTT GGG GAA GAC TTG CAA GCA ATG AGC CCT AA - #G GAA CTC CAG AAT CTA
 384
 Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
 115 - # 120 - # 125
 - - GAG CAA CAG CTT GAT ACT GCT CTT AAG CAC AT - #C CGC TCT AGA AAA AAC
 432
 Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Ser Arg Lys Asn
 130 - # 135 - # 140
 - - CAA CTT ATG TAC GAC TCC ATC AAT GAG CTC CA - #A AGA AAG GAG AAA GCC
 480
 Gln Leu Met Tyr Asp Ser Ile Asn Glu Leu Gl - #n Arg Lys Glu Lys Ala
 145 1 - #50 1 - #55 1 -
 #60
 - - ATA CAG GAA CAA AAC AGC ATG CTT TCC AAG CA - #G ATT AAG GAG AGG
 GAA 528
 Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg Glu
 165 - # 170 - # 175
 - - AAC GTT CTT AGG GCG CAA CAA GAG CAA TGG GA - #C GAG CAG AAC CAT GGC

576
Asn Val Leu Arg Ala Gln Gln Glu Gln Trp As - #p Glu Gln Asn His Gly
180 - # 185 - # 190
- - CAT AAT ATG CCT CCG CCT CCA CCC CCG CAG CA - #G CAT CAA ATC CAG CAT
624
His Asn Met Pro Pro Pro Pro Pro Pro Gln Gl - #n His Gln Ile Gln His
195 - # 200 - # 205
- - CCT TAC ATG CTC TCT CAT CAG CCA TCT CCT TT - #T CTC AAC ATG GGA GGG
672
Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
210 - # 215 - # 220
- - CTG TAT CAA GAA GAA GAT CAA ATG GCA ATG AG - #G AGG AAC GAT CTC GAT
720
Leu Tyr Gln Glu Glu Asp Gln Met Ala Met Ar - #g Arg Asn Asp Leu Asp
225 2 - #30 2 - #35 2 -
#40
- - CTG TCT CTT GAA CCC GTT TAC AAC TGC AAC CT - #T GGC CGT CGC TGC T
766
Leu Ser Leu Glu Pro Val Tyr Asn Cys Asn Le - #u Gly Arg Arg Cys
245 - # 250 - # 255
- - GA - # - # - #
768 - - - - (2) INFORMATION FOR SEQ ID NO:6:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 255 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
- - Met Gly Arg Gly Arg Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Met Lys Lys Ala
20 - # 25 - # 30
- - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Leu Val Val Phe
35 - # 40 - # 45
- - Ser His Lys Gly Lys Leu Phe Glu Tyr Pro Th - #r Asp Ser Cys Met Glu
50 - # 55 - # 60
- - Glu Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
65 - # 70 - # 75 - # 80
- - Ile Ala Pro Glu Ser Asp Ser Asn Thr Asn Tr - #p Ser Met Glu Tyr Asn
85 - # 90 - # 95
- - Arg Leu Lys Ala Lys Ile Glu Leu Leu Glu Ar - #g Asn Gln Arg His Tyr
100 - # 105 - # 110
- - Leu Gly Glu Asp Leu Gln Ala Met Ser Pro Ly - #s Glu Leu Gln Asn Leu
115 - # 120 - # 125
- - Glu Gln Gln Leu Asp Thr Ala Leu Lys His Il - #e Arg Ser Arg Lys Asn
130 - # 135 - # 140
- - Gln Leu Met Tyr Asp Ser Ile Asn Glu Leu Gl - #n Arg Lys Glu Lys Ala
145 1 - #50 1 - #55 1 -
#60
- - Ile Gln Glu Gln Asn Ser Met Leu Ser Lys Gl - #n Ile Lys Glu Arg
Glu 165 - # 170 - # 175
- - Asn Val Leu Arg Ala Gln Gln Glu Gln Trp As - #p Glu Gln Asn His Gly
180 - # 185 - # 190

- - His Asn Met Pro Pro Pro Pro Pro Gln Gl - #n His Gln Ile Gln His
 195 - # 200 - # 205
 - - Pro Tyr Met Leu Ser His Gln Pro Ser Pro Ph - #e Leu Asn Met Gly Gly
 210 - # 215 - # 220
 - - Leu Tyr Gln Glu Glu Asp Gln Met Ala Met Ar - #g Arg Asn Asp Leu Asp
 225 2 - #30 2 - #35 2 -
 #40
 - - Leu Ser Leu Glu Pro Val Tyr Asn Cys Asn Le - #u Gly Arg Arg Cys
 245 - # 250 - # 255
 - - - (2) INFORMATION FOR SEQ ID NO:7:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1345 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 149..968 - - (ix) FEATURE:
 (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..1345
 (D) OTHER INFORMATION: - #/note= "product = Zea mays AP1."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
 - - GCACGAGTCC TCCTCCTCCT CGCATCCAC CCCACCCAC CTTCTCCTTA AA -
 #GCTACCTG 60
 - - CCTACCCGGC GGTTCGCGC CGCAATCGAT CGACCGAAG AGAAAGAGCA GC -
 #TAGCTAGC 120
 - - TAGCAGATCG GAGCACGGCA ACAAGGCG ATG GGG CGC GGC AAG - #GTA CAG CTG
 172
 - # Met Gly Arg - #Gly Lys Val Gln Leu
 - # 1 - # 5
 - - AAG CGG ATA GAG AAC AAG ATA AAC CGG CAG GT - #G ACC TTC TCC AAG CGC
 220
 Lys Arg Ile Glu Asn Lys Ile Asn Arg Gln Va - #I Thr Phe Ser Lys Arg
 10 - # 15 - # 20
 - - CGG AAC GGC CTG CTC AAG AAG GCG CAC GAG AT - #C TCC GTC CTC TGC GAT
 268
 Arg Asn Gly Leu Leu Lys Lys Ala His Glu Il - #e Ser Val Leu Cys Asp
 25 - # 30 - # 35 - # 40
 - - GCC GAG GTC GCC GTC ATC GTC TTC TCC CCC AA - #G GGC AAG CTC TAC GAG
 316
 Ala Glu Val Ala Val Ile Val Phe Ser Pro Ly - #s Gly Lys Leu Tyr Glu
 45 - # 50 - # 55
 - - TAC GCC ACC GAC TCC CGC ATG GAC AAA ATT CT - #T GAA CGC TAT GAG CGA
 364
 Tyr Ala Thr Asp Ser Arg Met Asp Lys Ile Le - #u Glu Arg Tyr Glu Arg
 60 - # 65 - # 70
 - - TAT TCC TAT GCT GAA AAG GCT CTT ATT TCA GC - #T GAA TCT GAA AGT GAG
 412
 Tyr Ser Tyr Ala Glu Lys Ala Leu Ile Ser Al - #a Glu Ser Glu Ser Glu
 75 - # 80 - # 85
 - - GGA AAT TGG TGC CAC GAA TAC AGG AAA CTG AA - #G GCC AAA ATT GAG ACC
 460
 Gly Asn Trp Cys His Glu Tyr Arg Lys Leu Ly - #s Ala Lys Ile Glu Thr

90 - # 95 - # 100
 - - ATA CAA AAA TGC CAC AAG CAC CTG ATG GGA GA - #G GAT CTA GAG TCT TTG
 508
 Ile Gln Lys Cys His Lys His Leu Met Gly Gl - #u Asp Leu Glu Ser Leu
 105 1 - #10 1 - #15 1 -
 #20
 - - AAT CCC AAA GAG CTC CAG CAA CTA GAG CAG CA - #G CTG GAT AGC TCA
 CTG 556
 Asn Pro Lys Glu Leu Gln Gln Leu Glu Gln Gl - #n Leu Asp Ser Ser Leu
 10 125 - # 130 - # 135
 - - AAG CAC ATC AGA TCA AGG AAG AGC CAC CTT AT - #G GCC GAG TCT ATT TCT
 604
 Lys His Ile Arg Ser Arg Lys Ser His Leu Me - #t Ala Glu Ser Ile Ser
 140 - # 145 - # 150
 15 - - GAG CTA CAG AAG AAG GAG AGG TCA CTG CAG GA - #G GAG AAC AAG GCT CTG
 652
 Glu Leu Gln Lys Lys Glu Arg Ser Leu Gln Gl - #u Glu Asn Lys Ala Leu
 155 - # 160 - # 165
 - - CAG AAG GAA CTT GCG GAG AGG CAG AAG GCC GT - #C GCG AGC CGG CAG CAG
 700
 Gln Lys Glu Leu Ala Glu Arg Gln Lys Ala Va - #l Ala Ser Arg Gln Gln
 170 - # 175 - # 180
 - - CAG CAA CAG CAG CAG GTG CAG TGG GAC CAG CA - #G ACA CAT GCC CAG GCC
 748
 20 Gln Gln Gln Gln Gln Val Gln Trp Asp Gln Gl - #n Thr His Ala Gln Ala
 185 1 - #90 1 - #95 2 -
 #00
 - - CAG ACA AGC TCA TCA TCG TCC TCC TTC ATG AT - #G AGG CAG GAT CAG
 CAG 796
 Gln Thr Ser Ser Ser Ser Ser Ser Phe Met Me - #t Arg Gln Asp Gln Gln
 205 - # 210 - # 215
 - - GGA CTG CCG CCT CCA CAC AAC ATC TGC TTC CC - #G CCG TTG ACA ATG GGA
 844
 Gly Leu Pro Pro Pro His Asn Ile Cys Phe Pr - #o Pro Leu Thr Met Gly
 220 - # 225 - # 230
 - - GAT AGA GGT GAA GAG CTG GCT GCG GCG GCG GC - #G GCG CAG CAG CAG CAG
 892
 Asp Arg Gly Glu Glu Leu Ala Ala Ala Ala Al - #a Ala Gln Gln Gln Gln
 235 - # 240 - # 245
 40 - - CCA CTG CCG GGG CAG GCG CAA CCG CAG CTC CG - #C ATC GCA GGT CTG CCA
 940
 Pro Leu Pro Gly Gln Ala Gln Pro Gln Leu Ar - #g Ile Ala Gly Leu Pro
 250 - # 255 - # 260
 - - CCA TGG ATG CTG AGC CAC CTC AAT GCA T AAGG - #AGAGGG TCGATGAACA
 988 Pro Trp Met Leu Ser His Leu Asn Ala
 265 2 - #70
 - - CATCGACCTC CTCTCTCTCT CTCTCTCGTC ATGGATCATG ACGTACGCGT AC -
 #CATATGGT 1048
 - - TGCTGTGCCT GCGGCGATCG ATCGCGAGCA ATGGCAGCCT CATGCAAGTG AT -
 50 #CATTGCTC 1108
 - - CCCGTTGGTT AAACCCTAGC CTATGTTTAT GCGTCAGCA ACTAAGCTAA AC -

#TATTGTTA 1168
 - - TGTTCGCAAG AAAGGGTAAA CCCGCTAGCT GTGTAATCTT GTCCAGCTAT CA -
 #GTATGCTT 1228
 - - GTTACTGCCC AGTTACCCCTT GAATCTAGCG GCGCTTTTGG TGAGAGGGTG CA -
 #GTTTACTT 1288
 - - TAAACATGGT TCGTGACTTG CTGTAAATAG TAGTATTAAT CGATTTGGGC AT - #CTAAA
 1345 - - - (2) INFORMATION FOR SEQ ID NO:8:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 273 amino - #acids (B) TYPE: amino acid
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
 - - Met Gly Arg Gly Lys Val Gln Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Asn Gl - #y Leu Leu Lys Lys Ala
 15 20 - # 25 - # 30
 - - His Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ala Val Ile Val Phe
 35 - # 40 - # 45
 - - Ser Pro Lys Gly Lys Leu Tyr Glu Tyr Ala Th - #r Asp Ser Arg Met Asp
 50 - # 55 - # 60
 - - Lys Ile Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Ala Leu
 65 - # 70 - # 75 - # 80
 - - Ile Ser Ala Glu Ser Glu Ser Glu Gly Asn Tr - #p Cys His Glu Tyr Arg
 85 - # 90 - # 95
 - - Lys Leu Lys Ala Lys Ile Glu Thr Ile Gln Ly - #s Cys His Lys His Leu
 100 - # 105 - # 110
 - - Met Gly Glu Asp Leu Glu Ser Leu Asn Pro Ly - #s Glu Leu Gln Gln Leu
 115 - # 120 - # 125
 - - Glu Gln Gln Leu Asp Ser Ser Leu Lys His Il - #e Arg Ser Arg Lys Ser
 130 - # 135 - # 140
 - - His Leu Met Ala Glu Ser Ile Ser Glu Leu Gl - #n Lys Lys Glu Arg Ser
 145 1 - #50 1 - #55 1 -
 #60
 - - Leu Gln Glu Glu Asn Lys Ala Leu Gln Lys Gl - #u Leu Ala Glu Arg
 165 - # 170 - # 175
 - - Lys Ala Val Ala Ser Arg Gln Gln Gln Gln Gl - #n Gln Gln Val Gln Trp
 180 - # 185 - # 190
 - - Asp Gln Gln Thr His Ala Gln Ala Gln Thr Se - #r Ser Ser Ser Ser Ser
 195 - # 200 - # 205
 - - Phe Met Met Arg Gln Asp Gln Gln Gly Leu Pr - #o Pro Pro His Asn Ile
 210 - # 215 - # 220
 - - Cys Phe Pro Pro Leu Thr Met Gly Asp Arg Gl - #y Glu Glu Leu Ala Ala
 225 2 - #30 2 - #35 2 -
 #40
 - - Ala Ala Ala Ala Gln Gln Gln Gln Pro Leu Pr - #o Gly Gln Ala Gln
 245 - # 250 - # 255
 - - Gln Leu Arg Ile Ala Gly Leu Pro Pro Trp Me - #t Leu Ser His Leu Asn
 260 - # 265 - # 270 - - Ala
 - - - (2) INFORMATION FOR SEQ ID NO:9:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 779 base - #pairs

(B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 10..775 - - (ix) FEATURE:
 (A) NAME/KEY: unsure (B) LOCATION: 778..779
 (D) OTHER INFORMATION: - #/note= "N = one or more
 nucleotides. - #" - - (ix) FEATURE:
 (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..779
 (D) OTHER INFORMATION: - #/note= "product = Arabidopsis
 thaliana - #CAL."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
 - - TTAAGAGAA ATG GGA AGG GGT AGG GTT GAA TTG AAG - # AGG ATA GAG AAC
 48
 Met Gly Arg Gly Arg - #Val Glu Leu Lys Arg Ile Glu Asn
 1 - # 5 - # 10
 - - AAG ATC AAT AGA CAA GTG ACA TTC TCG AAA AG - #A AGA ACT GGT CTT TTG
 96
 Lys Ile Asn Arg Gln Val Thr Phe Ser Lys Ar - #g Arg Thr Gly Leu Leu
 15 - # 20 - # 25
 - - AAG AAA GCT CAG GAG ATC TCT GTT CTT TGT GA - #T GCC GAG GTT TCC CTT
 144
 Lys Lys Ala Gln Glu Ile Ser Val Leu Cys As - #p Ala Glu Val Ser Leu
 30 - # 35 - # 40 - # 45
 - - ATT GTC TTC TCC CAT AAG GGC AAA TTG TTC GA - #G TAC TCC TCT GAA TCT
 192
 Ile Val Phe Ser His Lys Gly Lys Leu Phe Gl - #u Tyr Ser Ser Glu Ser
 50 - # 55 - # 60
 - - TGC ATG GAG AAG GTA CTA GAA CGC TAC GAG AG - #G TAT TCT TAC GCC GAG
 240
 Cys Met Glu Lys Val Leu Glu Arg Tyr Glu Ar - #g Tyr Ser Tyr Ala Glu
 65 - # 70 - # 75
 - - AGA CAG CTG ATT GCA CCT GAC TCT CAC GTT AA - #T GCA CAG ACG AAC TGG
 288
 Arg Gln Leu Ile Ala Pro Asp Ser His Val As - #n Ala Gln Thr Asn Trp
 80 - # 85 - # 90
 - - TCA ATG GAG TAT AGC AGG CTT AAG GCC AAG AT - #T GAG CTT TTG GAG AGA
 336
 Ser Met Glu Tyr Ser Arg Leu Lys Ala Lys Il - #e Glu Leu Leu Glu Arg
 95 - # 100 - # 105
 - - AAC CAA AGG CAT TAT CTG GGA GAA GAG TTG GA - #A CCA ATG AGC CTC AAG
 384
 Asn Gln Arg His Tyr Leu Gly Glu Glu Leu Gl - #u Pro Met Ser Leu Lys
 110 1 - #15 1 - #20 1 -
 #25
 - - GAT CTC CAA AAT CTG GAG CAG CAG CTT GAG AC - #T GCT CTT AAG CAC
 ATT 432
 Asp Leu Gln Asn Leu Glu Gln Gln Leu Glu Th - #r Ala Leu Lys His Ile
 130 - # 135 - # 140
 - - CGC TCC AGA AAA AAT CAA CTC ATG AAT GAG TC - #C CTC AAC CAC CTC CAA
 480

Arg Ser Arg Lys Asn Gln Leu Met Asn Glu Se - #r Leu Asn His Leu Gln
 145 - # 150 - # 155
 - - AGA AAG GAG AAG GAG ATA CAG GAG GAA AAC AG - #C ATG CTT ACC AAA CAG
 528
 Arg Lys Glu Lys Glu Ile Gln Glu Glu Asn Se - #r Met Leu Thr Lys Gln
 160 - # 165 - # 170
 - - ATA AAG GAG AGG GAA AAC ATC CTA AAG ACA AA - #A CAA ACC CAA TGT GAG
 576
 Ile Lys Glu Arg Glu Asn Ile Leu Lys Thr Ly - #s Gln Thr Gln Cys Glu
 175 - # 180 - # 185
 - - CAG CTG AAC CGC AGC GTC GAC GAT GTA CCA CA - #G CCA CAA CCA TTT CAA
 624
 Gln Leu Asn Arg Ser Val Asp Asp Val Pro Gl - #n Pro Gln Pro Phe Gln
 190 1 - #95 2 - #00 2 -
 #05
 - - CAC CCC CAT CTT TAC ATG ATC GCT CAT CAG AC - #T TCT CCT TTC CTA
 AAT 672
 His Pro His Leu Tyr Met Ile Ala His Gln Th - #r Ser Pro Phe Leu Asn
 210 - # 215 - # 220
 - - ATG GGT GGT TTG TAC CAA GGA GAA GAC CAA AC - #G GCG ATG AGG AGG AAC
 720
 Met Gly Gly Leu Tyr Gln Gly Glu Asp Gln Th - #r Ala Met Arg Arg Asn
 225 - # 230 - # 235
 - - AAT CTG GAT CTG ACT CTT GAA CCC ATT TAC AA - #T TAC CTT GGC TGT TAC
 768
 Asn Leu Asp Leu Thr Leu Glu Pro Ile Tyr As - #n Tyr Leu Gly Cys Tyr
 240 - # 245 - # 250
 - - GCC GCT T GANN - # - # - # 779
 Ala Ala 255 - - - - (2) INFORMATION FOR SEQ ID NO:10:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 255 amino - #acids (B) TYPE: amino acid
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
 - - Met Gly Arg Gly Arg Val Glu Leu Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Thr Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - Gln Glu Ile Ser Val Leu Cys Asp Ala Glu Va - #l Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Val Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Arg Gln Leu
 65 - # 70 - # 75 - # 80
 - - Ile Ala Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Met Glu
 85 - # 90 - # 95
 - - Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Le - #u Glu Arg Asn Gln Arg
 100 - # 105 - # 110
 - - His Tyr Leu Gly Glu Glu Leu Glu Pro Met Se - #r Leu Lys Asp Leu Gln
 115 - # 120 - # 125
 - - Asn Leu Glu Gln Gln Leu Glu Thr Ala Leu Ly - #s His Ile Arg Ser Arg
 130 - # 135 - # 140

- - Lys Asn Gln Leu Met Asn Glu Ser Leu Asn Hi - #s Leu Gln Arg Lys Glu
 145 1 - #50 1 - #55 1 -
 #60
 - - Lys Glu Ile Gln Glu Glu Asn Ser Met Leu Th - #r Lys Gln Ile Lys
 5 165 - # 170 - # 175
 Glu
 - - Arg Glu Asn Ile Leu Lys Thr Lys Gln Thr Gl - #n Cys Glu Gln Leu Asn
 180 - # 185 - # 190
 - - Arg Ser Val Asp Asp Val Pro Gln Pro Gln Pr - #o Phe Gln His Pro His
 195 - # 200 - # 205
 10 - - Leu Tyr Met Ile Ala His Gln Thr Ser Pro Ph - #e Leu Asn Met Gly Gly
 210 - # 215 - # 220
 - - Leu Tyr Gln Gly Glu Asp Gln Thr Ala Met Ar - #g Arg Asn Asn Leu Asp
 225 2 - #30 2 - #35 2 -
 #40
 15 - - Leu Thr Leu Glu Pro Ile Tyr Asn Tyr Leu Gl - #y Cys Tyr Ala Ala
 245 - # 250 - # 255
 - - - - (2) INFORMATION FOR SEQ ID NO:11:
 - - (i) SEQUENCE CHARACTERISTICS:
 20 (A) LENGTH: 756 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 1..754 - - (ix) FEATURE:
 (A) NAME/KEY: misc.sub.-- - #feature
 25 (B) LOCATION: 1..756
 (D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
 CAL." - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
 - - ATG GGA AGG GGT AGG GTT GAA ATG AAG AGG AT - #A GAG AAC AAG ATC AAC
 48
 30 Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - CGA CAA GTG ACG TTT TCG AAA AGA AGA GCT GG - #T CTT TTG AAG AAA GCC
 96
 35 Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - CAT GAG ATC TCG ATC CTT TGT GAT GCT GAG GT - #T TCC CTT ATT GTC TTC
 144
 His Glu Ile Ser Ile Leu Cys Asp Ala Glu Va - #l Ser Leu Ile Val Phe
 35 - # 40 - # 45
 40 - - TCC CAT AAG GGG AAA CTG TTC GAG TAC TCG TC - #T GAA TCT TGC ATG GAG
 192
 Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - AAG GTA CTA GAA CAC TAC GAG AGG TAC TCT TA - #C GCC GAG AAA CAG CTA
 240
 45 Lys Val Leu Glu His Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu
 65 - # 70 - # 75 - # 80
 - - AAA GTT CCA GAC TCT CAC GTC AAT GCA CAA AC - #G AAC TGG TCA GTG GAA
 288
 50 Lys Val Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Val Glu
 85 - # 90 - # 95

- - TAT AGC AGG CTT AAG GCT AAG ATT GAG CTT TT - #G GAG AGA AAC CAA AGG
336

Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Le - #u Glu Arg Asn Gln Arg
100 - # 105 - # 110

- - CAT TAT CTG GGC GAA GAT TTA GAA TCA ATC AG - #C ATA AAG GAG CTA CAG
384

His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
115 - # 120 - # 125

- - AAT CTG GAG CAG CAG CTT GAC ACT TCT CTT AA - #A CAT ATT CGC TCG AGA
432

Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
130 - # 135 - # 140

- - AAA AAT CAA CTA ATG CAC GAG TCC CTC AAC CA - #C CTC CAA AGA AAG GAG
480

Lys Asn Gln Leu Met His Glu Ser Leu Asn Hi - #s Leu Gln Arg Lys Glu
145 1 - #50 1 - #55 1 -
#60

- - AAA GAA ATA CTG GAG GAA AAC AGC ATG CTT GC - #C AAA CAG ATA AGG
GAG 528

Lys Glu Ile Leu Glu Glu Asn Ser Met Leu Al - #a Lys Gln Ile Arg Glu
165 - # 170 - # 175

- - AGG GAG AGT ATC CTA AGG ACA CAT CAA AAC CA - #A TCA GAG CAG CAA AAC
576

Arg Glu Ser Ile Leu Arg Thr His Gln Asn Gl - #n Ser Glu Gln Gln Asn
180 - # 185 - # 190

- - CGC AGC CAC CAT GTA GCT CCT CAG CCG CAA CC - #G CAG TTA AAT CCT TAC
624

Arg Ser His His Val Ala Pro Gln Pro Gln Pr - #o Gln Leu Asn Pro Tyr
195 - # 200 - # 205

- - ATG GCA TCA TCT CCT TTC CTA AAT ATG GGT GG - #C ATG TAC CAA GGA GAA
672

Met Ala Ser Ser Pro Phe Leu Asn Met Gly Gl - #y Met Tyr Gln Gly Glu
210 - # 215 - # 220

- - TAT CCA ACG GCG GTG AGG AGG AAC CGT CTC GA - #T CTG ACT CTT GAA CCC
720

Tyr Pro Thr Ala Val Arg Arg Asn Arg Leu As - #p Leu Thr Leu Glu Pro
225 2 - #30 2 - #35 2 -

#40 - - ATT TAC AAC TGC AAC CTT GGT TAC TTT GCC GC - #A T GA
- # 756 Ile Tyr Asn Cys Asn Leu Gly Tyr Phe Ala Al - #a
245 - # 250

- - - (2) INFORMATION FOR SEQ ID NO:12:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 251 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear

- - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
- - Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile Asn
1 5 - # 10 - # 15
- - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
20 - # 25 - # 30

- - His Glu Ile Ser Ile Leu Cys Asp Ala Glu Val - #1 Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Val Leu Glu His Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu
 65 - # 70 - # 75 - # 80
 - - Lys Val Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Val Glu
 85 - # 90 - # 95
 - - Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Le - #u Glu Arg Asn Gln Arg
 100 - # 105 - # 110
 - - His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
 115 - # 120 - # 125
 - - Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
 130 - # 135 - # 140
 - - Lys Asn Gln Leu Met His Glu Ser Leu Asn Hi - #s Leu Gln Arg Lys Glu
 145 1 - #50 1 - #55 1 -
 #60
 - - Lys Glu Ile Leu Glu Glu Asn Ser Met Leu Al - #a Lys Gln Ile Arg
 Glu 165 - # 170 - # 175
 - - Arg Glu Ser Ile Leu Arg Thr His Gln Asn Gl - #n Ser Glu Gln Gln Asn
 180 - # 185 - # 190
 - - Arg Ser His His Val Ala Pro Gln Pro Gln Pr - #o Gln Leu Asn Pro Tyr
 195 - # 200 - # 205
 - - Met Ala Ser Ser Pro Phe Leu Asn Met Gly Gl - #y Met Tyr Gln Gly Glu
 210 - # 215 - # 220
 - - Tyr Pro Thr Ala Val Arg Arg Asn Arg Leu As - #p Leu Thr Leu Glu Pro
 225 2 - #30 2 - #35 2 -
 #40 - - Ile Tyr Asn Cys Asn Leu Gly Tyr Phe Ala Al - #a
 245 - # 250
 - - - (2) INFORMATION FOR SEQ ID NO:13:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 756 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
 - - (ix) FEATURE: (A) NAME/KEY: CDS
 (B) LOCATION: 1..451 - - (ix) FEATURE:
 (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..756
 (D) OTHER INFORMATION: - #/note= "product = Brassica oleracea
 var. botr - #ytis CAL."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
 - - ATG GGA AGG GGT AGG GTT GAA ATG AAG AGG AT - #A GAG AAC AAG ATC AAC
 48
 Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile Asn
 1 5 - # 10 - # 15
 - - AGA CAA GTG ACG TTT TCG AAA AGA AGA GCT GG - #T CTT TTG AAG AAA GCC
 96
 Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - CAT GAG ATC TCG ATT CTT TGT GAT GCT GAG GT - #T TCC CTT ATT GTC TTC
 144

His Glu Ile Ser Ile Leu Cys Asp Ala Glu Va - #1 Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - TCC CAT AAG GGG AAA CTG TTC GAG TAC TCG TC - #T GAA TCT TGC ATG GAG
 192
 5 Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - AAG GTA CTA GAA CGC TAC GAG AGG TAC TCT TA - #C GCC GAG AAA CAG CTA
 240
 10 Lys Val Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu
 65 - # 70 - # 75 - # 80
 - - AAA GCT CCA GAC TCT CAC GTC AAT GCA CAA AC - #G AAC TGG TCA ATG GAA
 288
 Lys Ala Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Met Glu
 85 - # 90 - # 95
 15 - - TAT AGC AGG CTT AAG GCT AAG ATT GAG CTT TG - #G GAG AGG AAC CAA AGG
 336
 Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Tr - #p Glu Arg Asn Gln Arg
 100 - # 105 - # 110
 - - CAT TAT CTG GGA GAA GAT TTA GAA TCA ATC AG - #C ATA AAG GAG CTA CAG
 384
 20 His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
 115 - # 120 - # 125
 - - AAT CTG GAG CAG CAG CTT GAC ACT TCT CTT AA - #A CAT ATT CGC TCC AGA
 432
 25 Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
 130 - # 135 - # 140
 - - AAA AAT CAA CTA ATG CAC T AGTCCCTCAA CCACCTCAA - #AGAAAGGAGA
 481 Lys Asn Gln Leu Met His 145 1 - #50
 - - AAGAAATACT GGAGGAAAAC AGCATGCTTG CCAAACAGAT AAAGGAGAGG GA -
 30 #GAGTATCC 541
 - - TAAGGACACA TCAAAACCAA TCAGAGCAGC AAAACCGCAG CCACCATGTA GC -
 #TCCTCAGC 601
 - - CGCAACCGCA GTTAAATCCT TACATGGCAT CATCTCCTTT CCTAAATATG GG -
 #TGGCATGT 661
 35 - - ACCAAGGAGA ATATCCAACG GCGGTGAGGA GGAACCGTCT CGATCTGACT CT -
 #TGAACCCA 721 - - TTTACAAC TG CAACCTTGGT TACTTTGCCG CATGA - #
 - # 756 - - - - (2) INFORMATION FOR SEQ ID NO:14:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 150 amino - #acids
 40 (B) TYPE: amino acid (D) TOPOLOGY: linear
 - - (ii) MOLECULE TYPE: protein
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
 - - Met Gly Arg Gly Arg Val Glu Met Lys Arg Il - #e Glu Asn Lys Ile
 Asn 1 5 - # 10 - # 15
 45 - - Arg Gln Val Thr Phe Ser Lys Arg Arg Ala Gl - #y Leu Leu Lys Lys Ala
 20 - # 25 - # 30
 - - His Glu Ile Ser Ile Leu Cys Asp Ala Glu Va - #1 Ser Leu Ile Val Phe
 35 - # 40 - # 45
 - - Ser His Lys Gly Lys Leu Phe Glu Tyr Ser Se - #r Glu Ser Cys Met Glu
 50 - # 55 - # 60
 - - Lys Val Leu Glu Arg Tyr Glu Arg Tyr Ser Ty - #r Ala Glu Lys Gln Leu


```

65          - # 70          - # 75          - # 80
- - Lys Ala Pro Asp Ser His Val Asn Ala Gln Th - #r Asn Trp Ser Met Glu
          85 - #          90 - #          95
- - Tyr Ser Arg Leu Lys Ala Lys Ile Glu Leu Tr - #p Glu Arg Asn Gln Arg
          100 - #          105 - #          110
- - His Tyr Leu Gly Glu Asp Leu Glu Ser Ile Se - #r Ile Lys Glu Leu Gln
          115 - #          120 - #          125
- - Asn Leu Glu Gln Gln Leu Asp Thr Ser Leu Ly - #s His Ile Arg Ser Arg
          130 - #          135 - #          140
- - Lys Asn Gln Leu Met His          145          1 - #50
- - - - (2) INFORMATION FOR SEQ ID NO:15:
- - (i) SEQUENCE CHARACTERISTICS:
- - (A) LENGTH: 1500 base - #pairs
- - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
- - (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: cDNA
- - (ix) FEATURE: (A) NAME/KEY: CDS
- - (B) LOCATION: 72..1343 - - (ix) FEATURE:
- - (A) NAME/KEY: misc.sub.-- - #feature
- - (B) LOCATION: 1..1500
- - (D) OTHER INFORMATION: - #/note= "product = Arabidopsis
          thaliana - #LEAFY (LFY). "
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
- - AAAGCAATCT GCTCAAAAGA GTAAAGAAAG AGAGAAAAAG AGAGTGATAG AG -
#AGAGAGAG          60
- - AAAAATAGAT T ATG GAT CCT GAA GGT TTC ACG AGT - #GGC TTA TTC CGG
TGG          110
          Met Asp Pro - #Glu Gly Phe Thr Ser Gly Leu Phe Arg Trp
          1 - #          5 - #          10
- - AAC CCA ACG AGA GCA TTG GTT CAA GCA CCA CC - #T CCG GTT CCA CCT CCG
          158
Asn Pro Thr Arg Ala Leu Val Gln Ala Pro Pr - #o Pro Val Pro Pro Pro
          15 - #          20 - #          25
- - CTG CAG CAA CAG CCG GTG ACA CCG CAG ACG GC - #T GCT TTT GGG ATG CGA
          206
Leu Gln Gln Gln Pro Val Thr Pro Gln Thr Al - #a Ala Phe Gly Met Arg
          30 - # 35 - # 40 - # 45
- - CTT GGT GGT TTA GAG GGA CTA TTC GGT CCA TA - #C GGT ATA CGT TTC TAC
          254
Leu Gly Gly Leu Glu Gly Leu Phe Gly Pro Ty - #r Gly Ile Arg Phe Tyr
          50 - #          55 - #          60
- - ACG GCG GCG AAG ATA GCG GAG TTA GGT TTT AC - #G GCG AGC ACG CTT GTG
          302
Thr Ala Ala Lys Ile Ala Glu Leu Gly Phe Th - #r Ala Ser Thr Leu Val
          65 - #          70 - #          75
- - GGT ATG AAG GAC GAG GAG CTT GAA GAG ATG AT - #G AAT AGT CTC TCT CAT
          350
Gly Met Lys Asp Glu Glu Leu Glu Glu Met Me - #t Asn Ser Leu Ser His
          80 - #          85 - #          90
- - ATC TTT CGT TGG GAG CTT CTT GTT GGT GAA CG - #G TAC GGT ATC AAA GCT
          398
Ile Phe Arg Trp Glu Leu Leu Val Gly Glu Ar - #g Tyr Gly Ile Lys Ala

```

20
 25
 30
 35
 40
 45
 50

95 - # 100 - # 105
 - - GCC GTT AGA GCT GAA CGG AGA CGA TTG CAA GA - #A GAG GAG GAA GAG GAA
 446
 Ala Val Arg Ala Glu Arg Arg Arg Leu Gln Gl - #u Glu Glu Glu Glu Glu
 110 1 - #15 1 - #20 1 -
 #25
 - - TCT TCT AGA CGC CGT CAT TTG CTA CTC TCC GC - #C GCT GGT GAT TCC
 GGT 494
 Ser Ser Arg Arg Arg His Leu Leu Leu Ser Al - #a Ala Gly Asp Ser Gly
 130 - # 135 - # 140
 - - ACT CAT CAC GCT CTT GAT GCT CTC TCC CAA GA - #A GAT GAT TGG ACA GGG
 542
 Thr His His Ala Leu Asp Ala Leu Ser Gln Gl - #u Asp Asp Trp Thr Gly
 145 - # 150 - # 155
 - - TTA TCT GAG GAA CCG GTG CAG CAA CAA GAC CA - #G ACT GAT GCG GCG GGG
 590
 Leu Ser Glu Glu Pro Val Gln Gln Gln Asp Gl - #n Thr Asp Ala Ala Gly
 160 - # 165 - # 170
 - - AAT AAC GGC GGA GGA GGA AGT GGT TAC TGG GA - #C GCA GGT CAA GGA AAG
 638
 Asn Asn Gly Gly Gly Gly Ser Gly Tyr Trp As - #p Ala Gly Gln Gly Lys
 175 - # 180 - # 185
 - - ATG AAG AAG CAA CAG CAG CAG AGA CGG AGA AA - #G AAA CCA ATG CTG ACG
 686
 Met Lys Lys Gln Gln Gln Gln Arg Arg Arg Ly - #s Lys Pro Met Leu Thr
 190 1 - #95 2 - #00 2 -
 #05
 - - TCA GTG GAA ACC GAC GAA GAC GTC AAC GAA GG - #T GAG GAT GAC GAC
 GGG 734
 Ser Val Glu Thr Asp Glu Asp Val Asn Glu Gl - #y Glu Asp Asp Asp Gly
 210 - # 215 - # 220
 - - ATG GAT AAC GGC AAC GGA GGT AGT GGT TTG GG - #G ACA GAG AGA CAG AGG
 782
 Met Asp Asn Gly Asn Gly Gly Ser Gly Leu Gl - #y Thr Glu Arg Gln Arg
 225 - # 230 - # 235
 - - GAG CAT CCG TTT ATC GTA ACG GAG CCT GGG GA - #A GTG GCA CGT GGC AAA
 830
 Glu His Pro Phe Ile Val Thr Glu Pro Gly Gl - #u Val Ala Arg Gly Lys
 240 - # 245 - # 250
 - - AAG AAC GGC TTA GAT TAT CTG TTC CAC TTG TA - #C GAA CAA TGC CGT GAG
 878
 Lys Asn Gly Leu Asp Tyr Leu Phe His Leu Ty - #r Glu Gln Cys Arg Glu
 255 - # 260 - # 265
 - - TTC CTT CTT CAG GTC CAG ACA ATT GCT AAA GA - #C CGT GGC GAA AAA TGC
 926
 Phe Leu Leu Gln Val Gln Thr Ile Ala Lys As - #p Arg Gly Glu Lys Cys
 270 2 - #75 2 - #80 2 -
 #85
 - - CCC ACC AAG GTG ACG AAC CAA GTA TTC AGG TA - #C GCG AAG AAA TCA
 GGA 974
 Pro Thr Lys Val Thr Asn Gln Val Phe Arg Ty - #r Ala Lys Lys Ser Gly

```

290 - # 295 - # 300
- - GCG AGT TAC ATA AAC AAG CCT AAA ATG CGA CA - #C TAC GTT CAC TGT TAC
1022
Ala Ser Tyr Ile Asn Lys Pro Lys Met Arg Hi - #s Tyr Val His Cys Tyr
5 305 - # 310 - # 315
- - GCT CTC CAC TGC CTA GAC GAA GAA GCT TCA AA - #T GCT CTC AGA AGA GCG
1070
Ala Leu His Cys Leu Asp Glu Glu Ala Ser As - #n Ala Leu Arg Arg Ala
320 - # 325 - # 330
10 - TTT AAA GAA CGC GGT GAG AAC GTT GGC TCA TG - #G CGT CAG GCT TGT TAC
1118
Phe Lys Glu Arg Gly Glu Asn Val Gly Ser Tr - #p Arg Gln Ala Cys Tyr
335 - # 340 - # 345
15 - AAG CCA CTT GTG AAC ATC GCT TGT CGT CAT GG - #C TGG GAT ATA GAC GCC
1166
Lys Pro Leu Val Asn Ile Ala Cys Arg His Gl - #y Trp Asp Ile Asp Ala
350 3 - #55 3 - #60 3 -
#65
- - GTC TTT AAC GCT CAT CCT CGT CTC TCT ATT TG - #G TAT GTT CCA ACA
20 AAG 1214
Val Phe Asn Ala His Pro Arg Leu Ser Ile Tr - #p Tyr Val Pro Thr Lys
370 - # 375 - # 380
- - CTG CGT CAG CTT TGC CAT TTG GAG CGG AAC AA - #T GCG GTT GCT GCG GCT
1262
25 Leu Arg Gln Leu Cys His Leu Glu Arg Asn As - #n Ala Val Ala Ala Ala
385 - # 390 - # 395
- - GCG GCT TTA GTT GGC GGT ATT AGC TGT ACC GG - #A TCG TCG ACG TCT GGA
1310
30 Ala Ala Leu Val Gly Gly Ile Ser Cys Thr Gl - #y Ser Ser Thr Ser Gly
400 - # 405 - # 410
- - CGT GGT GGA TGC GGC GGC GAC GAC TTG CGT TT - #C TAGTTTGGTT TGGGTAGTT
G 1363 Arg Gly Gly Cys Gly Gly Asp Asp Leu Arg Ph - #e
415 - # 420
35 - TGGTTGTTT AGTCGTTATC CTAATTAAC ATTAGTCTTT AATTTAGTCT TC -
#TTGGCTAA 1423
- - TTTATTTTTC TTTTGTGTC AAAACCTTTA ATTGTTATG GCTAATTTGT TA -
#TACACGCA 1483
- - GTTTTCTTAA TGCGTTA - # - #
40 - # 1500 - - - (2) INFORMATION FOR SEQ ID NO:16:
- - (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 424 amino - #acids (B) TYPE: amino acid
(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
45 - - Met Asp Pro Glu Gly Phe Thr Ser Gly Leu Ph - #e Arg Trp Asn Pro Thr
1 5 - # 10 - # 15
- - Arg Ala Leu Val Gln Ala Pro Pro Pro Val Pr - #o Pro Pro Leu Gln Gln
20 - # 25 - # 30
- - Gln Pro Val Thr Pro Gln Thr Ala Ala Phe Gl - #y Met Arg Leu Gly Gly
50 35 - # 40 - # 45
- - Leu Glu Gly Leu Phe Gly Pro Tyr Gly Ile Ar - #g Phe Tyr Thr Ala Ala

```

50 - # 55 - # 60
- - Lys Ile Ala Glu Leu Gly Phe Thr Ala Ser Th - #r Leu Val Gly Met Lys
65 - # 70 - # 75 - # 80
- - Asp Glu Glu Leu Glu Glu Met Met Asn Ser Le - #u Ser His Ile Phe Arg
5 85 - # 90 - # 95
- - Trp Glu Leu Leu Val Gly Glu Arg Tyr Gly Il - #e Lys Ala Ala Val Arg
100 - # 105 - # 110
- - Ala Glu Arg Arg Arg Leu Gln Glu Glu Gl - #u Glu Glu Ser Ser Arg
115 - # 120 - # 125
10 - - Arg Arg His Leu Leu Leu Ser Ala Ala Gly As - #p Ser Gly Thr His His
130 - # 135 - # 140
- - Ala Leu Asp Ala Leu Ser Gln Glu Asp Asp Tr - #p Thr Gly Leu Ser Glu
145 1 - #50 1 - #55 1 -
#60
15 - - Glu Pro Val Gln Gln Gln Asp Gln Thr Asp Al - #a Ala Gly Asn Asn
Gly 165 - # 170 - # 175
- - Gly Gly Gly Ser Gly Tyr Trp Asp Ala Gly Gl - #n Gly Lys Met Lys Lys
180 - # 185 - # 190
- - Gln Gln Gln Gln Arg Arg Arg Lys Lys Pro Me - #t Leu Thr Ser Val Glu
195 - # 200 - # 205
- - Thr Asp Glu Asp Val Asn Glu Gly Glu Asp As - #p Asp Gly Met Asp Asn
210 - # 215 - # 220
- - Gly Asn Gly Gly Ser Gly Leu Gly Thr Glu Ar - #g Gln Arg Glu His Pro
225 2 - #30 2 - #35 2 -
#40
- - Phe Ile Val Thr Glu Pro Gly Glu Val Ala Ar - #g Gly Lys Lys Asn
Gly 245 - # 250 - # 255
- - Leu Asp Tyr Leu Phe His Leu Tyr Glu Gln Cy - #s Arg Glu Phe Leu Leu
260 - # 265 - # 270
- - Gln Val Gln Thr Ile Ala Lys Asp Arg Gly Gl - #u Lys Cys Pro Thr Lys
275 - # 280 - # 285
- - Val Thr Asn Gln Val Phe Arg Tyr Ala Lys Ly - #s Ser Gly Ala Ser Tyr
290 - # 295 - # 300
- - Ile Asn Lys Pro Lys Met Arg His Tyr Val Hi - #s Cys Tyr Ala Leu His
305 3 - #10 3 - #15 3 -
#20
- - Cys Leu Asp Glu Glu Ala Ser Asn Ala Leu Ar - #g Arg Ala Phe Lys
Glu 325 - # 330 - # 335
- - Arg Gly Glu Asn Val Gly Ser Trp Arg Gln Al - #a Cys Tyr Lys Pro Leu
340 - # 345 - # 350
- - Val Asn Ile Ala Cys Arg His Gly Trp Asp Il - #e Asp Ala Val Phe Asn
355 - # 360 - # 365
- - Ala His Pro Arg Leu Ser Ile Trp Tyr Val Pr - #o Thr Lys Leu Arg Gln
370 - # 375 - # 380
45 - - Leu Cys His Leu Glu Arg Asn Asn Ala Val Al - #a Ala Ala Ala Ala Leu
385 3 - #90 3 - #95 4 -
#00
- - Val Gly Gly Ile Ser Cys Thr Gly Ser Ser Th - #r Ser Gly Arg Gly
Gly 405 - # 410 - # 415
50 - - Cys Gly Gly Asp Asp Leu Arg Phe 420
- - - (2) INFORMATION FOR SEQ ID NO:17:

```

- - (i) SEQUENCE CHARACTERISTICS:
    (A) LENGTH: 1656 base - #pairs
    (B) TYPE: nucleic acid (C) STRANDEDNESS: double
    (D) TOPOLOGY: linear - - (ix) FEATURE:
    (A) NAME/KEY: CDS (B) LOCATION: 1..1651
- - (ix) FEATURE: (A) NAME/KEY: misc.sub.-- - #feature
    (B) LOCATION: 1..1656
    (D) OTHER INFORMATION: - #/note= "domain = ecdysone receptor
        ligand bi - #nding domain."
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
- - ATG CCG CCG GAA TGC GTC GTC CCG GAG AAC CA - #A TGT GCG ATG AAG CCG
    48
Met Arg Pro Glu Cys Val Val Pro Glu Asn Gl - #n Cys Ala Met Lys Arg
    1      5 - #      10 - #      15
- - CGC GAA AAG AAG GCC CAG AAG GAG AAG GAC AA - #A ATG ACC ACT TCG CCG
    96
Arg Glu Lys Lys Ala Gln Lys Glu Lys Asp Ly - #s Met Thr Thr Ser Pro
    20 - #      25 - #      30
- - AGC TCT CAG CAT GGC GGC AAT GGC AGC TTG GC - #C TCT GGT GGC GGC CAA
    144
Ser Ser Gln His Gly Gly Asn Gly Ser Leu Al - #a Ser Gly Gly Gly Gln
    35 - #      40 - #      45
- - GAC TTT GTT AAG AAG GAG ATT CTT GAC CTT AT - #G ACA TGC GAG CCG CCC
    192
Asp Phe Val Lys Lys Glu Ile Leu Asp Leu Me - #t Thr Cys Glu Pro Pro
    50 - #      55 - #      60
- - CAG CAT GCC ACT ATT CCG CTA CTA CCT GAT GA - #A ATA TTG GCC AAG TGT
    240
Gln His Ala Thr Ile Pro Leu Leu Pro Asp Gl - #u Ile Leu Ala Lys Cys
    65 - # 70 - # 75 - # 80
- - CAA GCG CGC AAT ATA CCT TCC TTA ACG TAC AA - #T CAG TTG GCC GTT ATA
    288
Gln Ala Arg Asn Ile Pro Ser Leu Thr Tyr As - #n Gln Leu Ala Val Ile
    85 - #      90 - #      95
- - TAC AAG TTA ATT TGG TAC CAG GAT GGC TAT GA - #G CAG CCA TCT GAA GAG
    336
Tyr Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Gl - #u Gln Pro Ser Glu Glu
    100 - #      105 - #      110
- - GAT CTC AGG CGT ATA ATG AGT CAA CCC GAT GA - #G AAC GAG AGC CAA ACG
    384
Asp Leu Arg Arg Ile Met Ser Gln Pro Asp Gl - #u Asn Glu Ser Gln Thr
    115 - #      120 - #      125
- - GAC GTC AGC TTT CGG CAT ATA ACC GAG ATA AC - #C ATA CTC ACG GTC CAG
    432
Asp Val Ser Phe Arg His Ile Thr Glu Ile Th - #r Ile Leu Thr Val Gln
    130 - #      135 - #      140
- - TTG ATT GTT GAG TTT GCT AAA GGT CTA CCA GC - #G TTT ACA AAG ATA CCC
    480
Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Al - #a Phe Thr Lys Ile Pro
    145      1 - #50      1 - #55      1 -
#60

```

- - CAG GAG GAC CAG ATC ACG TTA CTA AAG GCC TG - #C TCG TCG GAG GTG
 ATG 528
 Gln Glu Asp Gln Ile Thr Leu Leu Lys Ala Cy - #s Ser Ser Glu Val Met
 165 - # 170 - # 175
 5 - - ATG CTG CGT ATG GCA CGA CGC TAT GAC CAC AG - #C TCG GAC TCA ATA TTC
 576
 Met Leu Arg Met Ala Arg Arg Tyr Asp His Se - #r Ser Asp Ser Ile Phe
 180 - # 185 - # 190
 10 - - TTC GCG AAT AAT AGA TCA TAT ACG CGG GAT TC - #T TAC AAA ATG GCC GGA
 624
 Phe Ala Asn Asn Arg Ser Tyr Thr Arg Asp Se - #r Tyr Lys Met Ala Gly
 195 - # 200 - # 205
 - - ATG GCT GAT AAC ATT GAA GAC CTG CTG CAT TT - #C TGC CGC CAA ATG TTC
 672
 15 - Met Ala Asp Asn Ile Glu Asp Leu Leu His Ph - #e Cys Arg Gln Met Phe
 210 - # 215 - # 220
 - - TCG ATG AAG GTG GAC AAC GTC GAA TAC GCG CT - #T CTC ACT GCC ATT GTG
 720
 Ser Met Lys Val Asp Asn Val Glu Tyr Ala Le - #u Leu Thr Ala Ile Val
 225 2 - #30 2 - #35 2 -
 #40
 - - ATC TTC TCG GAC CGG CCG GGC CTG GAG AAG GC - #C CAA CTA GTC GAA
 GCG 768
 Ile Phe Ser Asp Arg Pro Gly Leu Glu Lys Al - #a Gln Leu Val Glu Ala
 245 - # 250 - # 255
 25 - - ATC CAG AGC TAC TAC ATC GAC ACG CTA CGC AT - #T TAT ATA CTC AAC CGC
 816
 Ile Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Il - #e Tyr Ile Leu Asn Arg
 260 - # 265 - # 270
 30 - - CAC TGC GGC GAC TCA ATG AGC CTC GTC TTC TA - #C GCA AAG CTG CTC TCG
 864
 His Cys Gly Asp Ser Met Ser Leu Val Phe Ty - #r Ala Lys Leu Leu Ser
 275 - # 280 - # 285
 - - ATC CTC ACC GAG CTG CGT ACG CTG GGC AAC CA - #G AAC GCC GAG ATG TGT
 912
 35 - Ile Leu Thr Glu Leu Arg Thr Leu Gly Asn Gl - #n Asn Ala Glu Met Cys
 290 - # 295 - # 300
 - - TTC TCA CTA AAG CTC AAA AAC CGC AAA CTG CC - #C AAG TTC CTC GAG GAG
 960
 40 - Phe Ser Leu Lys Leu Lys Asn Arg Lys Leu Pr - #o Lys Phe Leu Glu Glu
 305 3 - #10 3 - #15 3 -
 #20
 - - ATC TGG GAC GTT CAT GCC ATC CCG CCA TCG GT - #C CAG TCG CAC CTT
 CAG 1008
 45 - Ile Trp Asp Val His Ala Ile Pro Pro Ser Va - #l Gln Ser His Leu Gln
 325 - # 330 - # 335
 - - ATT ACC CAG GAG GAG AAC GAG CGT CTC GAG CG - #G GCT GAG CGT ATG CGG
 1056
 50 - Ile Thr Gln Glu Glu Asn Glu Arg Leu Glu Ar - #g Ala Glu Arg Met Arg
 340 - # 345 - # 350
 - - GCA TCG GTT GGG GGC GCC ATT ACC GCC GGC AT - #T GAT TGC GAC TCT GCC

1104
Ala Ser Val Gly Gly Ala Ile Thr Ala Gly Il - #e Asp Cys Asp Ser Ala
355 - # 360 - # 365
- - TCC ACT TCG GCG GCG GCA GCC GCG GCC CAG CA - #T CAG CCT CAG CCT CAG
1152
Ser Thr Ser Ala Ala Ala Ala Ala Gln Hi - #s Gln Pro Gln Pro Gln
370 - # 375 - # 380
- - CCC CAG CCC CAA CCC TCC TCC CTG ACC CAG AA - #C GAT TCC CAG CAC CAG
1200
Pro Gln Pro Gln Pro Ser Ser Leu Thr Gln As - #n Asp Ser Gln His Gln
385 3 - #90 3 - #95 4 -
#00
- - ACA CAG CCG CAG CTA CAA CCT CAG CTA CCA CC - #T CAG CTG CAA GGT
CAA 1248
Thr Gln Pro Gln Leu Gln Pro Gln Leu Pro Pr - #o Gln Leu Gln Gly Gln
405 - # 410 - # 415
- - CTG CAA CCC CAG CTC CAA CCA CAG CTT CAG AC - #G CAA CTC CAG CCA CAG
1296
Leu Gln Pro Gln Leu Gln Pro Gln Leu Gln Th - #r Gln Leu Gln Pro Gln
420 - # 425 - # 430
- - ATT CAA CCA CAG CCA CAG CTC CTT CCC GTC TC - #C GCT CCC GTG CCC GCC
1344
Ile Gln Pro Gln Pro Gln Leu Leu Pro Val Se - #r Ala Pro Val Pro Ala
435 - # 440 - # 445
- - TCC GTA ACC GCA CCT GGT TCC TTG TCC GCG GT - #C AGT ACG AGC AGC GAA
1392
Ser Val Thr Ala Pro Gly Ser Leu Ser Ala Va - #l Ser Thr Ser Ser Glu
450 - # 455 - # 460
- - TAC ATG GGC GGA AGT GCG GCC ATA GGA CCC AT - #C ACG CCG GCA ACC ACC
1440
Tyr Met Gly Gly Ser Ala Ala Ile Gly Pro Il - #e Thr Pro Ala Thr Thr
465 4 - #70 4 - #75 4 -
#80
- - AGC AGT ATC ACG GCT GCC GTT ACC GCT AGC TC - #C ACC ACA TCA GCG
GTA 1488
Ser Ser Ile Thr Ala Ala Val Thr Ala Ser Se - #r Thr Thr Ser Ala Val
485 - # 490 - # 495
- - CCG ATG GGC AAC GGA GTT GGA GTC GGT GTT GG - #G GTG GGC GGC AAC GTC
1536
Pro Met Gly Asn Gly Val Gly Val Gly Val Gl - #y Val Gly Gly Asn Val
500 - # 505 - # 510
- - AGC ATG TAT GCG AAC GCC CAG ACG GCG ATG GC - #C TTG ATG GGT GTA GCC
1584
Ser Met Tyr Ala Asn Ala Gln Thr Ala Met Al - #a Leu Met Gly Val Ala
515 - # 520 - # 525
- - CTG CAT TCG CAC CAA GAG CAG CTT ATC GGG GG - #A GTG GCG GTT AAG TCG
1632
Leu His Ser His Gln Glu Gln Leu Ile Gly Gl - #y Val Ala Val Lys Ser
530 - # 535 - # 540
- - GAG CAC TCG ACG ACT GCA T AGCAG - # - #
1656 Glu His Ser Thr Thr Ala 545 5 - #50

- - - (2) INFORMATION FOR SEQ ID NO:18:

- - (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 550 amino - #acids (B) TYPE: amino acid

(D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

- Met Arg Pro Glu Cys Val Val Pro Glu Asn Gl - #n Cys Ala Met Lys Arg
1 5 - # 10 - # 15

- Arg Glu Lys Lys Ala Gln Lys Glu Lys Asp Ly - #s Met Thr Thr Ser Pro
20 - # 25 - # 30

- Ser Ser Gln His Gly Gly Asn Gly Ser Leu Al - #a Ser Gly Gly Gly Gln
35 - # 40 - # 45

- Asp Phe Val Lys Lys Glu Ile Leu Asp Leu Me - #t Thr Cys Glu Pro Pro
50 - # 55 - # 60

- Gln His Ala Thr Ile Pro Leu Leu Pro Asp Gl - #u Ile Leu Ala Lys Cys
65 - # 70 - # 75 - # 80

- Gln Ala Arg Asn Ile Pro Ser Leu Thr Tyr As - #n Gln Leu Ala Val Ile
85 - # 90 - # 95

- Tyr Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Gl - #u Gln Pro Ser Glu Glu
100 - # 105 - # 110

- Asp Leu Arg Arg Ile Met Ser Gln Pro Asp Gl - #u Asn Glu Ser Gln Thr
115 - # 120 - # 125

- Asp Val Ser Phe Arg His Ile Thr Glu Ile Th - #r Ile Leu Thr Val Gln
130 - # 135 - # 140

- Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Al - #a Phe Thr Lys Ile Pro
145 1 - #50 1 - #55 1 -
#60

- Gln Glu Asp Gln Ile Thr Leu Leu Lys Ala Cy - #s Ser Ser Glu Val
Met 165 - # 170 - # 175

- Met Leu Arg Met Ala Arg Arg Tyr Asp His Se - #r Ser Asp Ser Ile Phe
180 - # 185 - # 190

- Phe Ala Asn Asn Arg Ser Tyr Thr Arg Asp Se - #r Tyr Lys Met Ala Gly
195 - # 200 - # 205

- Met Ala Asp Asn Ile Glu Asp Leu Leu His Ph - #e Cys Arg Gln Met Phe
210 - # 215 - # 220

- Ser Met Lys Val Asp Asn Val Glu Tyr Ala Le - #u Leu Thr Ala Ile Val
225 2 - #30 2 - #35 2 -
#40

- Ile Phe Ser Asp Arg Pro Gly Leu Glu Lys Al - #a Gln Leu Val Glu
Ala 245 - # 250 - # 255

- Ile Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Il - #e Tyr Ile Leu Asn Arg
260 - # 265 - # 270

- His Cys Gly Asp Ser Met Ser Leu Val Phe Ty - #r Ala Lys Leu Leu Ser
275 - # 280 - # 285

- Ile Leu Thr Glu Leu Arg Thr Leu Gly Asn Gl - #n Asn Ala Glu Met Cys
290 - # 295 - # 300

- Phe Ser Leu Lys Leu Lys Asn Arg Lys Leu Pr - #o Lys Phe Leu Glu Glu
305 3 - #10 3 - #15 3 -
#20

- Ile Trp Asp Val His Ala Ile Pro Pro Ser Va - #l Gln Ser His Leu
Gln 325 - # 330 - # 335

- - Ile Thr Gln Glu Glu Asn Glu Arg Leu Glu Ar - #g Ala Glu Arg Met Arg
 340 - # 345 - # 350
 - - Ala Ser Val Gly Gly Ala Ile Thr Ala Gly Il - #e Asp Cys Asp Ser Ala
 355 - # 360 - # 365
 - - Ser Thr Ser Ala Ala Ala Ala Ala Gln Hi - #s Gln Pro Gln Pro Gln
 370 - # 375 - # 380
 - - Pro Gln Pro Gln Pro Ser Ser Leu Thr Gln As - #n Asp Ser Gln His Gln
 385 3 - #90 3 - #95 4 -
 #00
 - - Thr Gln Pro Gln Leu Gln Pro Gln Leu Pro Pr - #o Gln Leu Gln Gly
 Gln 405 - # 410 - # 415
 - - Leu Gln Pro Gln Leu Gln Pro Gln Leu Gln Th - #r Gln Leu Gln Pro Gln
 420 - # 425 - # 430
 - - Ile Gln Pro Gln Pro Gln Leu Leu Pro Val Se - #r Ala Pro Val Pro Ala
 435 - # 440 - # 445
 - - Ser Val Thr Ala Pro Gly Ser Leu Ser Ala Va - #l Ser Thr Ser Ser Glu
 450 - # 455 - # 460
 - - Tyr Met Gly Gly Ser Ala Ala Ile Gly Pro Il - #e Thr Pro Ala Thr Thr
 465 4 - #70 4 - #75 4 -
 #80
 - - Ser Ser Ile Thr Ala Ala Val Thr Ala Ser Se - #r Thr Thr Ser Ala
 Val 485 - # 490 - # 495
 - - Pro Met Gly Asn Gly Val Gly Val Gly Val Gl - #y Val Gly Gly Asn Val
 500 - # 505 - # 510
 - - Ser Met Tyr Ala Asn Ala Gln Thr Ala Met Al - #a Leu Met Gly Val Ala
 515 - # 520 - # 525
 - - Leu His Ser His Gln Glu Gln Leu Ile Gly Gl - #y Val Ala Val Lys Ser
 530 - # 535 - # 540
 - - Glu His Ser Thr Thr Ala 545 5 - #50
 - - - - (2) INFORMATION FOR SEQ ID NO:19:
 - - (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 855 base - #pairs
 (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear - - (ix) FEATURE:
 (A) NAME/KEY: CDS (B) LOCATION: 1..853
 - - (ix) FEATURE: (A) NAME/KEY: misc.sub.-- - #feature
 (B) LOCATION: 1..855
 (D) OTHER INFORMATION: - #/note= "domain = glucocorticoid
 receptor - #ligand binding domain."
 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
 - - ACA AAG AAA AAA ATC AAA GGG ATT CAG CAA GC - #C ACT GCA GGA GTC TCA
 48
 Thr Lys Lys Lys Ile Lys Gly Ile Gln Gln Al - #a Thr Ala Gly Val Ser
 1 5 - # 10 - # 15
 - - CAA GAC ACT TCG GAA AAT CCT AAC AAA ACA AT - #A GTT CCT GCA GCA TTA
 96
 Gln Asp Thr Ser Glu Asn Pro Asn Lys Thr Il - #e Val Pro Ala Ala Leu
 20 - # 25 - # 30
 - - CCA CAG CTC ACC CCT ACC TTG GTG TCA CTG CT - #G GAG GTG ATT GAA CCC
 144
 Pro Gln Leu Thr Pro Thr Leu Val Ser Leu Le - #u Glu Val Ile Glu Pro

35 - # 40 - # 45
 - - GAG GTG TTG TAT GCA GGA TAT GAT AGC TCT GT - #T CCA GAT TCA GCA TGG
 192
 Glu Val Leu Tyr Ala Gly Tyr Asp Ser Ser Va - #1 Pro Asp Ser Ala Trp
 50 - # 55 - # 60
 - - AGA ATT ATG ACC ACA CTC AAC ATG TTA GGT GG - #G CGT CAA GTG ATT GCA
 240
 Arg Ile Met Thr Thr Leu Asn Met Leu Gly Gl - #y Arg Gln Val Ile Ala
 65 - # 70 - # 75 - # 80
 - - GCA GTG AAA TGG GCA AAG GCG ATA CTA GGC TT - #G AGA AAC TTA CAC CTC
 288
 Ala Val Lys Trp Ala Lys Ala Ile Leu Gly Le - #u Arg Asn Leu His Leu
 85 - # 90 - # 95
 - - GAT GAC CAA ATG ACC CTG CTA CAG TAC TCA TG - #G ATG TTT CTC ATG GCA
 336
 Asp Asp Gln Met Thr Leu Leu Gln Tyr Ser Tr - #p Met Phe Leu Met Ala
 100 - # 105 - # 110
 - - TTT GCC TTG GGT TGG AGA TCA TAC AGA CAA TC - #A AGC GGA AAC CTG CTC
 384
 Phe Ala Leu Gly Trp Arg Ser Tyr Arg Gln Se - #r Ser Gly Asn Leu Leu
 115 - # 120 - # 125
 - - TGC TTT GCT CCT GAT CTG ATT ATT AAT GAG CA - #G AGA ATG TCT CTA CCC
 432
 Cys Phe Ala Pro Asp Leu Ile Ile Asn Glu Gl - #n Arg Met Ser Leu Pro
 130 - # 135 - # 140
 - - TGC ATG TAT GAC CAA TGT AAA CAC ATG CTG TT - #T GTC TCC TCT GAA TTA
 480
 Cys Met Tyr Asp Gln Cys Lys His Met Leu Ph - #e Val Ser Ser Glu Leu
 145 1 - #50 1 - #55 1 -
 #60
 - - CAA AGA TTG CAG GTA TCC TAT GAA GAG TAT CT - #C TGT ATG AAA ACC
 TTA 528
 Gln Arg Leu Gln Val Ser Tyr Glu Glu Tyr Le - #u Cys Met Lys Thr Leu
 165 - # 170 - # 175
 - - CTG CTT CTC TCC TCA GTT GCT AAG GAA GGT CT - #G AAG AGC CAA GAG TTA
 576
 Leu Leu Leu Ser Ser Val Ala Lys Glu Gly Le - #u Lys Ser Gln Glu Leu
 180 - # 185 - # 190
 - - TTT GAT GAG ATT CGA ATG ACT TAT ATC AAA GA - #G CTA GGA AAA GCC ATC
 624
 Phe Asp Glu Ile Arg Met Thr Tyr Ile Lys Gl - #u Leu Gly Lys Ala Ile
 195 - # 200 - # 205
 - - GTC AAA AGG GAA GGG AAC TCC AGT CAG AAC TG - #G CAA CGG TTT TAC CAA
 672
 Val Lys Arg Glu Gly Asn Ser Ser Gln Asn Tr - #p Gln Arg Phe Tyr Gln
 210 - # 215 - # 220
 - - CTG ACA AAG CTT CTG GAC TCC ATG CAT GAG GT - #G GTT GAG AAT CTC CTT
 720
 Leu Thr Lys Leu Leu Asp Ser Met His Glu Va - #1 Val Glu Asn Leu Leu
 225 2 - #30 2 - #35 2 -
 #40

```

- - ACC TAC TGC TTC CAG ACA TTT TTG GAT AAG AC - #C ATG AGT ATT GAA
TTC      768
Thr Tyr Cys Phe Gln Thr Phe Leu Asp Lys Th - #r Met Ser Ile Glu Phe
      245 - #      250 - #      255
5 - - CCA GAG ATG TTA GCT GAA ATC ATC ACT AAT CA - #G ATA CCA AAA TAT TCA
      816
Pro Glu Met Leu Ala Glu Ile Ile Thr Asn Gl - #n Ile Pro Lys Tyr Ser
      260 - #      265 - #      270
10 - - AAT GGA AAT ATC AAA AAG CTT CTG TTT CAT CA - #A AAA T GA
- #      855      Asn Gly Asn Ile Lys Lys Leu Leu Phe His Gl - #n Lys
      275 - #      280
- - - (2) INFORMATION FOR SEQ ID NO:20:
- - (i) SEQUENCE CHARACTERISTICS:
- - (A) LENGTH: 284 amino - #acids (B) TYPE: amino acid
15 - - (D) TOPOLOGY: linear - - (ii) MOLECULE TYPE: protein
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
- - Thr Lys Lys Lys Ile Lys Gly Ile Gln Gln Al - #a Thr Ala Gly Val Ser
1 5 - # 10 - # 15
- - Gln Asp Thr Ser Glu Asn Pro Asn Lys Thr Il - #e Val Pro Ala Ala Leu
20 - # 25 - # 30
- - Pro Gln Leu Thr Pro Thr Leu Val Ser Leu Le - #u Glu Val Ile Glu Pro
35 - # 40 - # 45
- - Glu Val Leu Tyr Ala Gly Tyr Asp Ser Ser Va - #l Pro Asp Ser Ala Trp
50 - # 55 - # 60
25 - - Arg Ile Met Thr Thr Leu Asn Met Leu Gly Gl - #y Arg Gln Val Ile Ala
65 - # 70 - # 75 - # 80
- - Ala Val Lys Trp Ala Lys Ala Ile Leu Gly Le - #u Arg Asn Leu His Leu
85 - # 90 - # 95
- - Asp Asp Gln Met Thr Leu Leu Gln Tyr Ser Tr - #p Met Phe Leu Met Ala
100 - # 105 - # 110
30 - - Phe Ala Leu Gly Trp Arg Ser Tyr Arg Gln Se - #r Ser Gly Asn Leu Leu
115 - # 120 - # 125
- - Cys Phe Ala Pro Asp Leu Ile Ile Asn Glu Gl - #n Arg Met Ser Leu Pro
130 - # 135 - # 140
35 - - Cys Met Tyr Asp Gln Cys Lys His Met Leu Ph - #e Val Ser Ser Glu Leu
145 1 - #50 1 - #55 1 -
#60
- - Gln Arg Leu Gln Val Ser Tyr Glu Glu Tyr Le - #u Cys Met Lys Thr
Leu 165 - # 170 - # 175
40 - - Leu Leu Leu Ser Ser Val Ala Lys Glu Gly Le - #u Lys Ser Gln Glu Leu
180 - # 185 - # 190
- - Phe Asp Glu Ile Arg Met Thr Tyr Ile Lys Gl - #u Leu Gly Lys Ala Ile
195 - # 200 - # 205
- - Val Lys Arg Glu Gly Asn Ser Ser Gln Asn Tr - #p Gln Arg Phe Tyr Gln
45 210 - # 215 - # 220
- - Leu Thr Lys Leu Leu Asp Ser Met His Glu Va - #l Val Glu Asn Leu Leu
225 2 - #30 2 - #35 2 -
#40
- - Thr Tyr Cys Phe Gln Thr Phe Leu Asp Lys Th - #r Met Ser Ile Glu
Phe 245 - # 250 - # 255
50 - - Pro Glu Met Leu Ala Glu Ile Ile Thr Asn Gl - #n Ile Pro Lys Tyr Ser

```

```

260      - #      265      - #      270
- - Asn Gly Asn Ile Lys Lys Leu Leu Phe His Gl - #n Lys
275      - #      280
- - - - (2) INFORMATION FOR SEQ ID NO:21:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 50 base - #pairs      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: double      (D) TOPOLOGY: linear
- - (ix) FEATURE:      (A) NAME/KEY: misc.sub.-- - #feature
      (B) LOCATION: 1..50
10      (D) OTHER INFORMATION: - #/note= "element = copper inducible
      regulatory - #element (ACE1 binding site)."
```

5

```

- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
- - AGCTTAGCGA TCGTCTTTT CCGCTGAACC GTTCCAGCAA AAAAGACTAG - #      50
- - - - (2) INFORMATION FOR SEQ ID NO:22:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 19 base - #pairs      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: double      (D) TOPOLOGY: linear
- - (ix) FEATURE:      (A) NAME/KEY: misc.sub.-- - #feature
      (B) LOCATION: 1..19
15      (D) OTHER INFORMATION: - #/note= "element = tet operator."
```

10

```

- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:
- - ACTCTATCAG TGATAGAGT - #      - #
- - # 19 - - - - (2) INFORMATION FOR SEQ ID NO:23:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 29 base - #pairs      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: double      (D) TOPOLOGY: linear
- - (ix) FEATURE:      (A) NAME/KEY: misc.sub.-- - #feature
      (B) LOCATION: 1..29
25      (D) OTHER INFORMATION: - #/note= "element = ecdysone response
      element."
```

20

```

- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
- - GATCCGACAA GGGTTCAATG CACTTGTC - #      - #
29 - - - - (2) INFORMATION FOR SEQ ID NO:24:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 371 base - #pairs
35      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: double      (D) TOPOLOGY: linear
- - (ix) FEATURE:      (A) NAME/KEY: misc.sub.-- - #feature
      (B) LOCATION: 1..371
40      (D) OTHER INFORMATION: - #/note= "element = heat shock
      inducible - #regulatory element (HSP81-1 promoter)."
```

30

```

- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
- - GTGGAGTCTC GAAACGAAAA GAACCTTCTG GAATTCGTTT GCTCACAAAG CT -
45 #AAAAACGG      60
- - TTGATTTTCAT CGAAATACGG CGTCGTTTTT AAAGAACAAT CCAGAAATCA CT -
#GGTTTTTC      120
- - TTTATTTCAA AAGAAGAGAC TAGAACTTTA TTTCTCCTCT ATAAATCAC TT -
#TGTTTTTC      180
50 - - CCTCTCTTCT TCATAAATCA ACAAACAAT CACAAATCTC TCGAAACGCT CT -
#CGAAGTTC      240
```

```

- - CAAATTTTCT CTTAGCATTC TCTTTCGTTT CTCGTTGCG TTGAATCAAA GT -
#TCGTTGCG      300
- - ATGGCGGATG TTCAGATGGC TGAIGCAGAG ACTTTTGCTT TCCAAGCTGA GA -
#TTAACCAG      360
5  - - CTTCTTAGCT T          - #          - #          - #          371
- - - - (2) INFORMATION FOR SEQ ID NO:25:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 29 base - #pairs          (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single              (D) TOPOLOGY: linear
10 - - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
- - GGATCCGGAT CAAAAATGGG AAGGGGTAG      - #          - #
      29 - - - - (2) INFORMATION FOR SEQ ID NO:26:
- - (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 30 base - #pairs          (B) TYPE: nucleic acid
15 - - (C) STRANDEDNESS: single              (D) TOPOLOGY: linear
- - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
- - GGATCCGCTG CGGCGAAGCA GCCAAGGTTG      - #          - #
      30

```

SEQ ID NO:27 and SEQ ID NO:28

Arabidopsis SEP1 cDNA and Arabidopsis SEP1 amino acid sequence

```

      20      40      60      80      100
ATGGGAAGAG GAAGAGTAGA GCTGAAGAGG ATAGAGAACA AAATCAACAG ACAAGTAACG TTGCAAAAGC GTAGGAACGG TTTGTTGAAG AAAGCTTATG
TACCCCTTCTC CTCTCATCTC CGACTTCTCC TATCTCTTGT TTTAGTTGTC TGTTCATTGC AAACGTTTCG CATCCTTGCC AAACAACCTC TTTGGAATAC
M G R G R V E L K R I E N K I N R Q V T F A K R R N G L L K K A Y>

      120      140      160      180      200
AATTGTCTGT TCTCTGTGAT GCTGAAGTTG CTCTCATCAT CTCTCCAAC CGTGGAAAGC TCTATGAGTT TTGCAGCTCC TCAACATATG TCAAGACACT
TTAACAGACA AGAGACACTA CGACTTCAAC GAGAGTAGTA GAAGAGGTTG GCACCTTTTC AGATACTCAA AACGTCGAGG AGTTTGTACG AGTTCTGTGA
E L S V L C D A E V A L I I F S N R G K L Y E F C S S S N M L K T L>

      220      240      260      280      300
TGATCGGTAC CAGAAATGCA GCTATGGATC CATTGAAGTC AACAACAACG TCCGCAAGA ACTTGAGAAC AGCTACAGAG AATATCTGAA GCTTAAGGGT
ACTAGCCATG GTCTTTACGT CGATACCTAG GTAACCTCAG TTGTTGTTTG GACGGTTTCT TGAACCTCTG TCGATGCTC TTATAGACTT CGAATTCCTA
D R Y Q K C S Y G S I E V N N K P A K E L E N S Y R E Y L K L K G>

      320      340      360      380      400
AGATATGAGA ACCTTCAACG TCAACAGAGA AATCTTCTTG GGGAGGATTT AGGACCTTTG AATTCAAAGG AGTTAGAGCA GCTTGAGCGT CAACTGGACG
TCTATACTCT TGAAGTTGCG AGTTGTCTCT TTAGAAGAAC CCTCTCTAAA TCCTGGAAC TTAAGTTTCC TCAATCTCGT CGAATCTGCA GTTGACCTGC
R Y E N L Q R Q Q R N L L G E D L G P L N S K E L E Q L E R Q L D>

      420      440      460      480      500
GCTCTCTCAA GCAAGTTCGG TCCATCAAGA CACAGTACAT GCTTGACCAG CTCTCGGATC TTCAAAATAA AGAGCAAATG TTGCTTGAAA CCAATAGAGC
CGAGAGAGTT CGTTCAAGCC AGGTAGTTCT GTGTCTATGTA CGAAGTGGTC GAGAGCCTAG AAGTTTATT TCTCGTTTAC AACGAACCTT GGTTATCTCG
G S L K Q V R S I K T Q Y M L D Q L S D L Q N K E Q M L L E T N R A>

      520      540      560      580      600
TTTGGCAATG AAGCTGGATG ATATGATTGG TGTGAGAAGT CATCATATGG GAGGATGGGA AGGCGGTGAA CAGAATGTTA CCTACGCGCA TCATCAAGCT
AAACCGTTAC TTGACCTTAC TATACTAACC ACACTCTTCA GTAGTATAAC CTCTACCTCT TCCGCCACTT GTCTTACAAT GGATGCGCGT AGTATGTCGA
L A M K L D D M I G V R S H H M G G W E G G E Q N V T Y A H H Q A>

      620      640      660      680      700
CAGTCTCAGG GACTATACCA GCCTCTTGAA TGCAATCCAA CTCTGCAART GGGGTATGAT AATCCAGTAT GCTCTGAGCA AATCACTGCG ACAACACAAG
GTCAGAGTCC CTGATATGGT CGGAGAACCT ACCTTAGGTT GAGAGGTTTA CCCCACTATA TTAGGTCATA CGAGACTCGT TTAGTGACGC TGTTGTGTTT
Q S O G L Y Q P L E C N P T L Q M G Y D N P V C S E Q I T A T T Q>

      720
CTCAGGCGCA GCGGGGAAC GGTACATTC CAGGATGGAT GCTCTGA
GAGTCCGCGT CGGCCCTTTG CCAATGTAAG GTCCTACCTA CGAGACT
A Q A Q P G N G Y I P G W M L *>

```

SEQ ID NO:29 and SEQ ID NO:30

Arabidopsis SEP2 cDNA and Arabidopsis SEP2 amino acid sequence

```

20          40          60          80          100
ATGGAAGAG GAAGAGTAGA GCTCAAGAGG ATAGAGAACA AAATCAACAG ACAAGTGACG TTTGCTAAAC GTAGAAATGG TTTGCTGAAA AAAGCTTATG
TACCCTTCTC CTCTCTATCT CGAGTTCTCC TATCTCTTGT TTTAGTTGTC TGTTCACTGC AAACGATTTG CATCTTTACC AAACGACTTT TTTGGAATAC
M G R G R V E L K R I E N K I N R Q V T F A K R R N G L L K K A Y>

120          140          160          180          200
AGCTTTCTGT TCTCTCGGAT GCTGAAGTCT CTCTCATCGT CTCTCCAAC CGTGGCAAGC TCTACGAGTT CTGCAGCACC TCCAACATGC TCAAGACACT
TCGAAAGACA AGAGACGCTA CGACTTCAGA GAGAGTAGCA GAAGAGGTTG GCACCGTTCC AGATGCTCAA GACGTCGTGG AGGTTGTACG AGTTCTGTGA
E L S V L C D A E V S L I V F S N R G K L Y E F C S T S N M L K T L>

220          240          260          280          300
GGAAAGGTAT CAGAAGTGT GCTATGGCTC CATTGAAGTC AACAAACAAAC CTGCTAAAGA GCTTGAGAAC AGCTACAGAG AGTACTTGAA GCTGAAAGGT
CCTTTCCATA GTCTTCCATC CATACCGAG GTAACCTCAG TTGTTGTTTG GACGATTCTT CGAACTCTTG TCGATGTCTC TCATGAACCT CGACTTTCCA
E R Y Q K C S Y G S I E V N N K P A K E L E N S Y R E Y L X L K G>

320          340          360          380          400
AGATATGAAA ATCTGCAACG TCAGCAGAGA AATCTTCTTG GAGAGGATCT TGGACCTCTG AATTCAAAGG AGCTAGAGCA GCTTGAGCGT CAACTAGACG
TCTATACTTT TAGACGTTGC AGTCGTCTCT TTAGAAGAAC CTCTCCTAGA ACCTGGAGAC TTAAGTTTCC TCGATCTCGT CGAACTCGCA GTTGATCTGC
R Y E N L Q R Q Q R N L L G E D L G P L N S K E L E Q L E R Q L D>

420          440          460          480          500
GCTCTCTGAA GCAAGTTGCG TGCATCAAGA CACAGTATAT GCTTGACCAG CTCTCTGATC TTCAAGGTAA GGAGCATATC TTGCTTGATG CCAACAGAGC
CGAGAGACTT CGTTCAAGCG ACGTAGTTCT GTGTCATATA CGAACTGGTC GAGAGACTAG AAGTTCCATT CCTCGTATAG AACGAACATC GGTGTTCTCG
G S L K Q V R C I K T Q Y M L D Q L S D L Q G K E H I L L D A N R A>

520          540          560          580          600
TTTGTCATG AAGCTGGAAG ATATGATCGG CGTGAGACAT CACCATATAG GAGGAGGATG GGAAGGTGGT GATCAACAGA ATATTGCCTA TGGACATCCT
AAACAGTTAC TTGACCTTTC TATACTAGCC GCACTCTGTA GTGGTATATC CTCTCTCTAC CCTTCCACCA CTAGTTGTCT TATAACGGAT ACCTGTAGGA
L S M K L E D M I G V R H H H I G G G W E G G D Q Q N I A Y G H P>

620          640          660          680          700
CAGGCTCATT CTCAGGGACT ATACCAATCT CTTGAATGTG ATCCCACTTT GCAAATTGGA TATAGCCATC CAGTGTGCTC AGAGCAAATG GCTGTGACGG
GTCCGAGTAA GAGTCCCTGA TATGTTAGA GAACCTACAC TAGGGTGAAA CGTTTAACTT ATATCGGTAG GTACACAGAG TCTCTGTTAC CGACACTGCC
Q A H S Q G L Y Q S L E C D P T L Q I G Y S H P V C S E Q M A V T>

720          740
TGCAAGGTCA GTCCCAACAA GGAAACGGCT ACATCCCTGG CTGGATGCTG TGA
ACGTTCCAGT CAGGGTTGTT CCTTTGCGA TGTAGGGACC GACCTACGAC ACT
V Q G Q S Q Q G N G Y I P G W M L *>

```

SEQ ID NO:31 and SEQ ID NO:32

Arabidopsis SEP3 cDNA and Arabidopsis SEP3 amino acid sequence

```

20          40          60          80          100
ATGGAAGAG GGAGAGTAGA ATTGAAGAGG ATAGAGAACA AGATCAATAG GCAAGTGACG TTTGCAAGA GAAGGAATGG TCTTTTGAAG AAAGCATATG
TACCCTTCTC CCTCTCATCT TAACITCTCC TATCTCTTGT TCTAGTTATC CGTTCACTGC AAACGTTTCT CTTCCTTACC AGAAAACCTC TTTGCTATGC
M G R G R V E L K R I E N K I N R Q V T F A K R R N G L L K K A Y>

120          140          160          180          200
AGCTTTCACT TCTATGTGAT GCAGAAGTTG CTCTCATCAT CTCTCAAAT AGAGGAAAGC TGTACGAGTT TTGACGATGT TCGAGCATGC TCGGACACT
TCGAAAGTCA AGATACACTA CGTCTTCAAC GAGAGTAGTA GAAGAGTTTA TCTCTTTCCG ACATGCTCAA AACGTCATCA AGCTCGTACG AACCTGTGTA
E L S V L C D A E V A L I I F S N R G K L Y E F C S S S S M L R T L>

220          240          260          280          300
GGAGAGGTAC CAAAAGTGTA ACTATGGAGC ACCAGAACCC AATGTGCTT CAAGAGAGGC CTTAGCAGTT GAACTTAGTA GCCAGCAGGA GTATCTCAAG
CCTCTCCATG GTTTTCACAT TGATACCTCG TGGTCTTGGG TTACACGGAA GTTCTCTCCG GAATCGTCAA CTTGAATCAT CGGTCTGCTC CATAGAGTTC
E R Y Q K C N Y G A P E P N V P S R E A L A V E L S S Q Q E Y L K>

320          340          360          380          400
CTTAAGGAGC GTTATGACGC CTTACAAAGA ACCCAAGGA ATCTGTTGGG AGAAGATCTT GGACCTCTAA GTACAAAGGA GCTTGAGTCA CTTGAGAGAC
GAATCTCTCG CAATACCTCG GAATGTTTCT TGGGTTTCTT TAGACAACC TCTCTAGAA CTTGAGATTT CATGTTTCTC CGAACTCAGT GAACTCTCTG
L K E R Y D A L Q R T Q R N L L G E D L G P L S T K E L E S L E R>

420          440          460          480          500
AGCTTGATTC TTCTCTGAAG CAGATCAGAG CTCTCAGGAC ACAGTTTATG CTTGACCAGC TCAACGATCT TCAGAGTAAG TTAGCTGATG GGTATCAGAT
TCGAACCTAG AAGGAACCTC GTCTAGTCTC GAGAGTCTCG TGTCAAATAC GAACTGGTCC AGTTGCTAGA AGTCTCATTC AATCGACTAC CCATAGTCTA

```

Q L D S S L K Q I R A L R T Q P M L D Q L N D L Q S K L A D G Y Q M>

5
10
15
520 540 560 580 600
GCCACTCCAG CTGAACCCCTA ACCAAGAAGA GGTGTGATCAC TACGGTGCSTC ATCATCATCA ACAACAACAA CACTCCCAAG CTCTTCTCCA GCCTTTGGAA
CGGTGAGGTC GACTTGGGAT TGGTTCCTCT CCAACTAGTG ATGCCAGCAG TAGTAGTAGT TGTGTGTGTT GTGAGGGTTC GAAAGAAAGT CGGAACCTT
P L Q L N P N Q E E V D H Y G R H H H Q Q Q Q H S Q A F F Q P L E>
620 640 660 680 700
TGTGAACCCA TTCTTCAGAT CGGGTATCAG GGGCAGCAAG ATGGAATGGG AGCAGGACCA AGTGTGAATA ATTACATGTT GGGTTGGTTA CCTTATGACA
ACACTGGGT AAGAAGCTTA GCCCATAGTC CCGTCGTTT TACCTTACCC TCGTCCTGGT TCACACTTAT TAATGTACAA CCCAACCAAT GGAATACTGT
C E P I L Q I G Y Q G Q Q D G M G A G P S V N N Y M L G W L P Y D>
CCAACTCTAT TTGA
GGTTGAGATA AACT
T N S I *>

SEQ ID NO:33 and SEQ ID NO:34

Arabidopsis AGL20 cDNA and Arabidopsis AGL20 amino acid sequence

20
25
30
35
40
45
50
20 40 60 80 100
ATGTTGAGGG GCAAACTCA GATGAAGAGA ATAGAGAATG CAACAAGCAG ACAAGTGACT TTCTCCAAA GAAGGAATGG TTTGTGAAG AAAGCCTTTG
TACCCTCCG CTTTGTGAT CTACTCTCT TATCTCTTAC GTTGTTCGTG TGTTCACCTGA AAGAGGTTTT CTTCCTTACC AAACAACCTT TTTCCGAAAC
M V R G K T Q M K R I E N A T S R Q V T F S K R R N G L L K K A F>
120 140 160 180 200
AGCTCTCAGT GCTTGTGAT GCTGAAGTTT CTCTTATCAT CTCTCTCTCT AAAGGCAAA TTTATGAAT CGCCAGCTCC AATATGCAAG ATACCATAGA
TCGAGATCA CGAAACACTA CGACTTCAAA GAGAATAGTA GAAGAGAGGA TTTCCGTTTG AAATACTTAA CGCGTCGAGG TTATACGTTT TATGGTATCT
E L S V L C D A E V S L I I F S P K G R L Y E F A S S N M Q D T I D>
220 240 260 280 300
TCGTTATCTG AGGCATACCTA AGGATCGAGT CAGCACCAAA CCGGTTTCTG AAGAAAAATAT GCAGCATTTG AAATATGAAG CAGCAAAACAT GATGAAGAAA
AGCAATAGAC TCCGTATGAT TCCTAGCTCA GTCGTGGTTT GGCCAAAGAC TCTTTTATA CGTCGTAAAC TTTTACTTTC GTCGTTTGTA CTACTCTTTT
R Y L R H T K D R V S T K P V S E E N M Q H L K Y E A A N M M K K>
320 340 360 380 400
ATTGAACAAC TCGAAGCTTC TAAACGTAAA CTCTTGGGAG AAGGCATAGG AACATGCTCA ATCGAGGAGC TGCAACAGAT TGAGCAACAG CTTGAGAAAA
TAACTTGTG AGCTTCGAAG ATTTCATTT GAGAACCCTC TTCCGTATCC TTGTACGAGT TAGCTCCTCG ACGTTGTCTA ACTCGTTGTC GAACTCTTTT
I E Q L E A S K R K L L G E G I G T C S I E E L Q Q I E Q Q L E K>
420 440 460 480 500
GTGTCAAATG TATTGAGCA AGAAAGACTC AAGTCTTTAA GGAACAAAT GAGCAGCTCA AGCAAAAGGA GAAAGCTCTA CTTGCAGAAA ACGAGAAAGCT
CACAGTTTAC ATAAGCTCGT TCTTTCTGAG TTCACAAAT CTTTGTTTAA CTCGTCGAGT TCGTTTCTCT CTTCGAGAT CGAGCTCTTT TGCTCTCTGA
S V K C I R A R K T Q V P K E Q I E Q L K Q K E K A L A A E N E K L>
520 540 560 580 600
CTCTGAAAAG TGGGGATCTC ATGAAGCGA AGTTTGGTCA AATAAGAATC AAGAAAGTAC TGGAAGAGGT GATGAAGAGA GTAGCCCAAG TTCTGAAGTA
GAGACTTTTC ACCCTAGAG TACTTTCGCT TCAAACAGT TTAATCTTAG TTCTTTCATG ACCCTCTCCA CTACTCTCTC CATCGGGTTC AAGACTTCAT
S E K W G S H E S E V W S N K N Q E S T G R G D E E S S P S S E V>
620 640
GAGACGCAAT TGTTCATTGG GTTACCTTGT TCTTCAAGAA AGTGA
CTCTGCGTTA ACAAGTAACC CAATGGAACA AGAAGTTCTT TCACT
E T Q L F I G L P C S S R K *>

SEQ ID NO:35 and SEQ ID NO:36

Arabidopsis AGL22 cDNA and Arabidopsis AGL22 amino acid sequence

55
60
65
20 40 60 80 100
ATGGCGAGAG AAAAGATTCA GATCAGGAAG ATCGACAACG CAACGGCGAG ACAAGTGAGC TTTTCGAAAC GAAGAAGAGG GCTTTTCAAG AAAGCTGAAG
TACCGCTCTC TTTTCTAAGT CTAGTCTTTC TAGCTGTTCG GTTGCCGCTC TGTTCACCTC AAAAGCTTTG CTTCCTTCTC CGAAAAGTTC TTTGCACTTC
M A R E K I Q I R K I D N A T A R Q V T F S K R R R G L F K K A E>
120 140 160 180 200
AACTCTCGT TCTCTGCGAC GCCGATGTCG CTCTCATCAT CTCTCTTCTC ACCGGAAAAC TGTTGAGATT CTGTAGCTCC AGCATGAAGG AAGTCTTAGA
TTGAGAGGCA AGAGACCGTG CGGCTACAGC GAGAGTAGTA GAAGAGAAGG TGGCCTTTTG ACAAGCTCAA GACATCGAGG TCGTACTTCC TTCGATGCT
E L S V L C D A D V A L I I F S S T G K L F E F C S S S M K E V L E>
220 240 260 280 300

GAGGCATAAC TTTCAGTCAA AGAACTTGGG GAAGCTTCAT CAGCCATCTC TTGAGTTACA GCTGTTGAG AACAGTGATC ACGCCCGAAT GAGTAAAGAA
CTCCGTATTG AANGTCAGTT TCTTGAACCT CTTCGAAGTA GTCGGTAGAG AACTCAATGT CGACCAACTC TTGTCAGTAG TCGCGGCTTA CTCATTCTTT
R H N X Q S K N L E K L H Q P S L E L Q L V E N S D H A R M S K E>

5

320 340 360 380 400
ATTGCGGACA AGAGCCACCG ACTAAGGCAA ATGAGAGGAG AGGAACTTCA AGGACTTGAC ATTGAAGAGC TTGACAGCT AGAGAAGGCC CTTGAAACTG
TAACGCCTGT TCTCGTGGC TGATTCGGTT TACTCTCTCT TCCTTGAAGT TCCTGAACTG TAACCTCTCG AAGTGTGCGA TCTCTCCGG GAACCTTGAC
I A D K S H R L R Q M R G E E L Q G L D I E E L Q Q L E K A L E T>

10

420 440 460 480 500
GTTTGACGCG TGTGATTGAA ACAAGAGAGT ACAAGATTAT GAGTGAGATC AGCGAACTTC AGAAAAAGGG AATGCAATTG ATGGATGAGA ACAAGCGGTT
CAAACTGCGC ACACAACTTT TGTTTCTCAC TGTTCTAATA CTCACCTTAG TCGCTTGAAG TCTTTTCCC TTACGTTAAC TACCTACTCT TGTTGCGCAA
G L T R V I E T K S D K I M S E I S E L Q K K G M Q L M D E N K R L>

15

520 540 560 580 600
GAGGCAGCAA GTATGTGTCT TACCCTCTCT GTTGATAACA AATCCCTTTC TTTTGTCTAC CATTAAAGTA CACACTCTTA AATTAAATCC CCAGTTGTCT
CTCCGTGCTT CATACACAGA ATGGGAGAGA CAACTATTGT TTAGGGAAAG AAAACAGATG GTAATTGCAT GTGTGAGGAT TTAATATTAGG GGTCAACAGA
R Q Q V C V L P S L L I T N P F L L S T I N V H T P K F N P Q L S>

20

620
ACAACACATA TGTTTGATCA TACTGTGAGA TAA
TGTTGTGTAT ACAAACTAGT ATGACACTCT ATT
T T H M F D H T V R *>

25

SEQ ID NO:37 and SEQ ID NO:38

Arabidopsis AGL24 cDNA and Arabidopsis AGL24 amino acid sequence

30

20 40 60 80 100
ATGCGGAGAG AGAAGATAAG GATAAAGAAG ATTGATAACA TAACAGCGAG ACAAGTTACT TTCTCAAAGA GAAGAAGAGG AATCTTCAAG AAAGCGGATG
TACCGCTCTC TCTTCTATTC CTATTCTTTC TAACATTGTG ATTGTCGCTC TGTTCAATGA AAGAGTTTCT CTCTTCTCC TTAGAAGTTC TTTCGGCTAC
M A R E K I R I K K I D N I T A R Q V T F S K R R R G I F K K A D>

35

120 140 160 180 200
AACTTTCAGT TCTTTCGGAT GCTGATGTTG CTCTCATCAT CTCTCTGCCC ACCGGAAGGC TCTTCGAGTT CTCCAGCTCA AGAATGAGAG ACATATTGGG
TTGAAAGTCA AGAAACGCTA CGACTACAAC GAGAGTAGTA GAAGAGACGG TGCCCTTTCC AGAAGCTCAA GAGGTGAGAT TCTTACTCTC TGTATAACCC
E L S V L C D A D V A L I I F S A T G K L F E F S S S R M R D I L G>

40

220 240 260 280 300
AAGGTATAGT CTTCATGCAA GTAACATCAA CAAATTGATG GATCCACCTT CTACTCATCT CCGGCTTGAG AATTGTAACC TCTCCAGACT AAGTAAGGAA
TTCCATATCA GAAGTACGTT CATTTAGTGT GTTTAACTAC CTAGGTGGAA GATGAGTAGA GGCCGAAGCT TTAACATTGG AGAGGTCTGA TTCATTCTTT
R Y S L H A S N I N K L M D P P S T H L R L E N C N L S R L S K E>

45

320 340 360 380 400
GTCGAAGACA AAACCAAGCA GCTACGGAAA CTGAGAGGAG AGGATCTTGA TGGATTGAAC TTAGAAGAGT TGCAGCGGCT GGAGAACTA CTTGAATCCG
CAGCTTCTGT TTTGTTCTGT CGATGCCCTT GACTCTCTCT TCCTAGAAGT ACCTAACTTG AATCTTCTCA ACCTGCGCGA CCTCTTTGAT GAACCTAGGC
V E D K T K Q L R K L R G E D L D G L N L E E L Q R L E K L L E S>

50

420 440 460 480 500
GACTTAGCCG TGTGTCTGAA AAGAAGGGCG AGTGTGTGAT GAGCCAAATT TTCTCACTTG AGAAACGGGG ATCGGAATTG GTGGATGAGA ATAAGAGACT
CTGAATCGGC ACACAGACTT TTCTTCCCGC TCACACACTA CTCGGTTTAA AAGAGTGAAC TCTTTGCCCC TAGCCTTAAC CACCTACTCT TATTCTGTGA
G L S R V S E K K G E C V M S Q I F S L E K R G S E L V D E N K R L>

55

520 540 560 580 600
GAGGGATAAA CTAGAGACGT TGGAAAGGGC AAAACTGAGC ACGCTTAAAG AGCCTTTGGA GACAGAGTCG GTGACACAA ATGTGTCAAG CTACGACAGT
CTCCCTATTG GATCTCTGCA ACCTTTCCCG TTTTGACTGC TCGGAATTTC TCCGAAACCT CTGTCTCAGC CACTGGTGTG TACACAGTTC GATGCTGTCA
R D K L E T L E R A K L T T L K E A L E T E S V T T N V S S Y D S>

60

620 640 660
GGAACTCCCC TTGAGGATGA CTCCGACACT TCCTTGAAGC TTGGGCTTCC ATCTTGGGAA TGA
CCTTGAGGGG AACTCTACT GAGGCTGTGA AGGACTTTCG AACCGAAGG TAGAACCCTT ACT
G T P L E D D S D T S L K L G L P S W E *>

SEQ ID NO:39 and SEQ ID NO:40

Arabidopsis AGL27 cDNA and Arabidopsis AGL27 amino acid sequence

65

20 40 60 80 100
ATGGGAAGAA GAAAAATCGA GATCAAGCGA ATCGAGAGCA AAAGCAGTCG ACAAGTCACT TTCTCCAAAC GACGCAATGG TCTCATCGAC AAAGCTCGAC
TACCCCTCTT CTCTTTAGCT CTAGTTCGCT TAGCTCTTGT TTTCGTCAGC TGTTCAAGTA AAGAGGTTTG CTCGCTTACC AGAGTAGCTG TTTCGAGCTG

M G R R K I E I K R I E N K S S R Q V T F S K R R N G L I D K A R >

120 140 160 180 200
AACTTTCGAT TCTCTGTGAA TCCTCCGTCG CTGTGTGCGT CGTATCTGCC TCCGGAACAC TCTATGACTC TTCCTCCGGT GACGACATTT CCAAGATCAT
TTGAAAGCTA AGAGACACTT AGGAGGCAGC GACAACAGCA GCATAGACGG AGGCCCTTTG AGATACTGAG AAGGAGGCCA CTGCTGTAAA GGTTCCTAGTA
Q L S I L C E S S V A V V V V S A S G K L Y D S S S G D D I S K I I >

220 240 260 280 300
TGATCGTTAT GAAATACAAC ATGCTGATGA ACTTAGAGCC TTAGATCTTG AAGAAAAAAT TCAGAATTAT CTCCACACA AGGAGTTACT AGAAACAGTC
ACTAGCAATA CTTTATGTTG TACGACTACT TGAATCTCGG AATCTAGAAC TTCTTTTTTA AGTCTTAATA GAAGGTGTGT TCCTCAATGA TCTTTGTCAG
D R Y E I Q H A D E L R A L D L E E K I Q N Y L P H K E L L E T V >

320 340 360 380 400
CAAAGCAAGC TTGAAGAACC AAATGTCGAT AATGTAAGTG TAGATTCTCT AATTTCCTCG GAGGAACAAC TTGAGACTGC TCTGTCCGTA AGTAGAGCTA
GTTTCGTTCC AACTTCTTGG TTTACAGCTA TTACATTAC ATCTAAGAGA TTAAGAGAGC CTCCTTGTGT AACTCTGACG AGACAGGCAT TCATCTCGAT
Q S K L E E P N V D N V S V D S L I S L E E Q L E T A L S V S R A >

420 440 460 480 500
GGAAGGCAGA ACTGATGATG GAGTATATCG AGTCCCTTAA AGAAAAGGAG AAATTGCTGA GAGAAGAGAA CCAGGTCTCG GCTAGCCAGC TGTGAGAGAA
CTTCCGTCCT TGACTACTAC CTCATATAGC TCAGGGAATT TCTTTTCCTC TTAAACGACT CTCTTCTCTT GGTCCAAGAC CGATCGGTGG ACAGTCTCTT
R K A E L M M E Y I E S L K E K E K L L R E E N Q V L A S Q L S E K >

520 540 560 580 600
GAAAGGTATG TCTCACCGAT GAAAGATACT CAAACCCGGA TGGGAAGGAA TACGTTGCTG GCAACAGATG ATGAGAGAGG AATGTTTCCG GGAAGTAGCT
CTTTCATAC AGAGTGGCTA CTTTCTATGA GTTTTGGGCT ACCCTTTCTT ATGCAACGAC CGTGTCTTAC TACTCTCTCC TTACAAGGCG CCTTCATCGA
K G M S H R * K I L K T R W E R I R C W Q Q M M R E E C F R E V A >

620 640 660 680
CCGGCAACAA AATACCGGAG ACTCTCCCGC TGCTCAATTA GCCACCATCA TCAACGGCTG AGTTTTCACC TTAACCTCAA AGCCTGA
GGCCGTTGTT TTATGGCCTC TGAGAGGGCG ACCAGTTAAT CGGTGGTAGT AGTTGCCGAC TCAAAGTGG AATTGTAGTT TCGGACT
P A T K Y R R L S R C S I S H H H Q R L S F H L K L K A >

SEQ ID NO:41

Arabidopsis SEP1 genomic sequence

-2981 -2961 -2941 -2921 -2901
CAGATCTCTT GGCATGTGTC GAAATGTGG AGATCTTAAG AATGTAGCTT GTGGCGGTG CAAAGGAACA GGAACAATCA AATCAGGAGG ATTCTTTGGT
GTCTAGAGAA CCGTACACAG CTTTACACG TCTAGAATTC TTACATCGAA CACCGGCAAC GTTTCCTTGT CTTTGTTAGT TTAGTCTCTC TAAGAAACCA

-2881 -2861 -2841 -2821 -2801
TTCAGTGACT CATCAACAC AAGTACAGTG GCTTGGGATA ATTGCCAAGC CAAAGGTGTG TTCCCTTGCC CTGAATGCTC AAAATCTTGA CCATTTTCTC
AAGTCACTGA GTAGTTTGTG TTCTAGTCA CCAACGCTAT TAACGGTTCG GTTTCACACA AAGGGAACCG GACTTACGAG TTTTAGAAGT GGTAAAGAG

-2781 -2761 -2741 -2721 -2701
GGTATTTTAT AGTTTGTTC TCTTCTGAC ACTATGATAA GTGTAATCGG TCCATTGGTA ATGGTAATGT TAAAGTTGAA GAATGTCTTG TTTATTGAG
CCATAAAATA TCAACAAAGT AGAAGAACTG TGATACTATT CACATTAGCC AGGTAACCAT TACCATTACA ATTTCACCTT CTACACAGAA AAATAAGCTC

-2681 -2661 -2641 -2621 -2601
AAGTCTCTTA TTCCAATTCT TGATCTGTTA CTGCAATATA GGCACCTTTC TTAGATGTAC CGGATGCTTA TGAATTACTG AGTAGGTTAA CTTTAAACCG
TTCAGAGAAT AAGGTTAAGA ACTAGACAAAT GACGTTTATT CCGTGAAGAC AATCTACATG GCCTACGAAT ACTTAATGAC TCATCCAATT GAAATGGGCC

-2581 -2561 -2541 -2521 -2501
GTTTTATCGT CATTAAACCG GAGAAATTC TCTAGTAACC AAATGCTCTG CTGGACCTTT CTTTCAGTGA GCAACTATAG GTGGGTTTTT GGCAGTTGAT
CAAAATAGCA GTAAATTTGGC CTCTTTAAGT AGATCAATTG TTTACGAGAC GACCTGGAAA GAAAGTCACT CGTTGATATC CACCCAAAAA CCGTCAACTA

-2481 -2461 -2441 -2421 -2401
GTACCATAAT TGGTGCAAAAC ACACATTTTT CTTGAATTTT TGTTTAACTT AAATAAAGTT ACTTCGTTTT CTTGTTTTTT TTAATATGAA TAAAAAAAT
CATGGTATTA ACCACGTTTG TGTGTAATAA GAACCTTAAA ACAAAATTGAA TTTATTTCAA TGAAGCAAAA GAACAAAAAA AATTATACCT ATTTTTTTTA

-2381 -2361 -2341 -2321 -2301
CAACCATAAC TGATAGTAGG TTGGTTATCT TTATCAAAAC AAATAAAGTT AATAGGCAGA AAAATAATTG TCTATAGAAAT CAATTATGAA AATGCCATTT
GTTGGTATTG ACTATCATCC AACCAATAGA AATAGTTTGT TTTATTTCAA TTATCCGTCT TTTTATTAAC AGATATCTTA GTTAATACCT TTACGGTAAA

-2281 -2261 -2241 -2221 -2201
TTTGGGATGG CATTTGTGGA TTTTGCCCTT TTTTAAATAG TTTGTGAATT TTGCCATTTT TCAGGTTACG TGAATGAATA TACGTTTTAT TCATTATGTT
AAACCTTACC GTAAACACCT AAAACGGGAA AAAAATTATC AAACACTTAA AACGGTAAAA AGTCCAATGC ACTTACTTAT ATGCAAAATA AGTAATACAA

-2181 -2161 -2141 -2121 -2101

TGGGTTTACT CGGTTGGT TGTCTTAGG GTTTAGTATT TTGTGTAAC TACGTATTTT TACCAAAAA AGTCGGAAT CCATATATTT TTAATCTTA
 ACCCAAATGA GCCAACACCA ACAAGAATCC CAAATCATAA AACACATTTG ATGCATAAAA ATGGTTTTTT TCAGGCTTTT GGTATATAAA AATTITAGAA

-2081 -2061 -2041 -2021 -2001
 GAAAATGGCT TATCCGTAAG ATTTTAGTAA AAATGGCAAT TTCAAAGAT CTCTATAAAA AATGGCAAAA TCAACAATAA TCCCTTGTCT ATATGGTGGT
 CTTTTACCGA ATAGGCATTC TAAATCATTT TTTACCGTTA AAGTTTTCTA GAGATATTTT TTACCGTTTT AGTTGTATT AGGGAACAGA TATACCACCA

-1981 -1961 -1941 -1921 -1901
 ATTTCTGCTA AAAAGTACTT ATGGGTAGAT TTTTAGCTT CATAGATTC TTGTGAAAA AAAATTACTT TGACATTTT AGTGGAGTTA TTTAAATTTT
 TAAAGACGAT TTTCACTGAA TACCATCTA AAAATCGAA GTATCTAAGA AACAGCTTTT TTTTAAATGAA ACATGTAAAA TCACCTCAAT AATTTTAAAG

-1881 -1861 -1841 -1821 -1801
 CCAATGAAC AAAACCATAT ATTGATGAAA TTGCAAAATG CAATCCAAAA ATAAATATGT TCCACTCTTT TGGTTAGCTT TTAACATAAG ATCGGTTTTA
 GGTTAACTTG TTTTGGTATA TAACACTTTT AAGCGTTTAC GTTAGGTTTT TATTTATACA AGGTGAGAAA ACCAATCGAA AATTGATTTT TACGAAAAAT

-1781 -1761 -1741 -1721 -1701
 CTTTATGTA GTGGTTGATC TTTTGGCAAT GGGGACAAAT GACTATACAA TCTAAGAGAT CATTTTAAAG AATATCATTC ATATTTCATC CTCTCTTCA
 GAAATACATT CACCAACTAG AAAACCGTTA CCCCTGTGTA CTGATATGTT AGATTTCTTA GTAAATTTGC TTATAGTAAG TATAAAGTAG GAGAGAAGT

-1681 -1661 -1641 -1621 -1601
 AATTTTCAGTT TCACTAATTA ACCACGTTTC AATTGTAGTG TATCGCGAGC TGTAATATTT ATCTAATTTA TGTTCATATA TCATAACTGT AATCTTTTAT
 TTAAGTCAA AGTGATTAAT TGGTGCAAG TTAACATCAC ATAGCGCTCG ACATTTATAA TAGATTAAAT ACAATGTATT AGTATTGACA TTAGAAATAA

-1581 -1561 -1541 -1521 -1501
 AGACAAAAAC ATATATACCT CACTGCAAC ACCTTCAAC ATGGATAACT TGATTTAGGC ATACAAATAT TATTTCTCAT TTATTTGATA TGACCTATAT
 TCTGTTTTTG TATATATGGA GTGACGTTTG TGGAGTTTG TACCTATTGA ACTAATCCG TATGTTTTATA ATAAAGATA AATAAACTAT ACTGGATATA

-1481 -1461 -1441 -1421 -1401
 TATGTGGCTA TTTTATCAGT TTTAGTGTIT TTTATGATAA TTGAACCACT TAAATGTTA TCTCATTTTT CAATTTATTT TAAACTGAAT TAAAAAGTAA
 ATACACCGAT AAAATAGTCA AAATCACAAA AAATACTATT AACTTGGTGA ATTTACAAAT AGAGTAAAAA GTTAAATAAA ATTTGACTTA ATTTTTCATT

-1381 -1361 -1341 -1321 -1301
 GAAAGTATGA TCCAATAAGG CATCGACACA TGGAAACCCA TTTTAAAGTA GAAGATGCTT TTCTGCGGCT TCTGAAAAA ACTAGAAAAA GATATGATAC
 CTTTCATACT AGGTATTTCC GTAGCTGTGT ACCTTTGGGT AAAATTTCCAT CTCTACGAA AAGACGCCGA AGACTTTTTG TGATCTTTTA CTATACTATG

-1281 -1261 -1241 -1221 -1201
 GTTGCTTTCA TTATATGTAA GTATTATTTA GTTTTAAATC ACGCGCTTCA TATCCAGCTG CAAGACTACT ACAACTTGCA ATTATGAGAC TCTGTTAGA
 CAACGAAAGT AATAACATT CATATAAAT CAAATTAAG TGGCGGAAGT ATAGGTGAGT GTTCTGATGA TGTGAACGT TAATATCTGT AGAGCAATCT

-1181 -1161 -1141 -1121 -1101
 AAATTACCAG GTATAATTTA AAAACAAAAA GAACTAGAAT ATATTGGCAA TTATTGGAAG TAAGAAAAATA TGAGATTCTT GACCGAGTTG TTAACATATC
 TTTAATGGTC CATATTAAT TTTTGTTTTT CTGATCTTTA TATAACCGTT AATAAACTTC ATCTTTTAT ACTCTAAGAA CTGGCTCAAC AATTTGATAG

-1081 -1061 -1041 -1021 -1001
 AAACCCAAAA GTTTTGGTTA AAAAATAAGC TAGTACTATG TACATATGTT TTATGTTGAA AATATATTAA ACTGTATGTA AGAGGGAGTG TACTTTTCAAT
 TTTGGTTTTT CAAACCAAT TTTTATTTTG ATCATGATAC ATGTATACAA AATACAACTT TTATATAATT TGACATACAT TCTCCCTCAC ATGAAAGTAA

-981 -961 -941 -921 -901
 TTAGATATAC ATTTCCAGCT AGTACGAGGT CTCTATATAT AAACTTTCTT AATATCGCTA AACAAATTTT ACTTTCAAGT TTGTATGTTG ATAAAGTAAA
 AATCTATATG TAAAGTGCGA TCATGCTCCA GAGATATATA TTTGAAAGAA TTATAGCGAT TTGTTTAAAA TGAAAGTTCA AACATTACAC TATTCACITT

-881 -861 -841 -821 -801
 GACCGTATAT ACATACACAT GTTAATCAAC TGATAACCTT TGTGCTCGT GTGTCTAGTT ACTAGTCAAC CATCAAAGT GCATGATGCT GTTTTCTTA
 CTGSCATATA TGTATGTCTA CAATTAGTTG ACTAATGSA ACGGAGCA CACAGATCAA TGATCAGTTG GTAGTTTGCA CGTACTACGA CAAAAGAAAT

-781 -761 -741 -721 -701
 GAGTACTATT GTTGTGTTAT ATATAACTAA ACATAAACAA TTTGCTATTA TGATATAAAC ATAGAATTTT CAAGCAATGA TATGTTTGA TGTTTTGTAT
 CTCTAGTAAA CAACACAATA TATATTGATT TGTAATTTGT AAACGATAAT ACTATATTG TATCTTAAAA GTTCTGTACT ATACAAATCT ACAAACATA

-681 -661 -641 -621 -601
 AAATATTCCA TAAATAGTAG ACACCCATAT ATACACAAC ATGAATTTCTA CCTGAGGAGA AACACATAGA TGTTCAAAT AATAATAAC CCTATAATGA
 TTTATAAGGT ATTTATCATC TGTGGGTATA TATGTGTTTG TACTTAAGAT GGACTCCTCT TTGTGTATCT ACAAGTTTAA TTTATTTATG GGATATTACT

-581 -561 -541 -521 -501
 AAACCTTAAA GTAAGTAATA CGAAATAAAA ATTTATCTTT TAAATAACAT ATAAACATAT ATATACAAGT TTAATGGTGA ATTTGATCAC AAGAGCCAAT
 TTTGAGATTT CATTCATTAT GCTTTATTTT TAAATAGGAA ATTTATGTA TATTTGTATA TATATGTTCA AATTAACCAT TAACATAGTG TTCTCGGTTA

-481 -461 -441 -421 -401

TATTGGTGA CTGTATCACA CGTGCTTAAA GAGAGCGTGG GAATGAAAGT AAAGAAGAAT AAAGAAGCAG AGAGATGGGC TAGAAATGAG AAAACACACC
ATAAACCACT GACATAGTGT GCACGAATTT CTCTCGCACC CTACTTTCA TTTCTTCTTA TTTCTTCGTC TCTCTACCCG ATCTTTACTC TTTTGTGTGG

-381 -361 -341 -321 -301
AAACCCTAAC CTCACCCCTCA CACATTTCTT ATCTTTTGCT CTCAATAGAT TCCATTGATT CAAACACAAA TTTTCATTAA GATTTACAAA CCTCCACACA
TTTGGGATTG GAGTGGGAGT GTGTAAAGAA TAGAAAACGA GAGTTATCTA AGGTAACATA GTTTTGTTTT AAAAGTAATT CTAAAGTGTT GGAGTTGTGT

-281 -261 -241 -221 -201
CTTCCAAACA CAATTAAGAA GAGGAAAAAG AATCAATAAC CCTATAATA AAAAATCAGA CAAACAGAAG TTTCTCTTTC TTCTTCTTTC AGCTAGTAGC
GAAGGTTTGT GTTAATTTCT CTCTTTTTC TTAGTTATTG GGATATTTAT TTTTGTAGCT GTTTGTCTTC AAAGGAGAAG AAGAAGGAAT TCGATCATGG

-181 -161 -141 -121 -101
TTTTGTCTCT GAAATTAGGG TTAATTTCTT TTTTCCAAT ACCATCAATT CTCCAGACCA TAAAACTCA AAAAGATCAG ATCTTTCCTC TGAAAAAGAG
AAAAACAAGAA CTTTAATCCC AATTAAAGAA AAAAGGTTTA TGTAGTTAA GAGGTCTGGT ATTTTGTAGT TTTTCTAGTC TAGAAAGGAG ACTTTTCTCT

-81 -61 -41 -21 -1
ATACCCAACT TATGTTTTG TGTGCTCTGA TATAGATAAA CATTACATAC CCATATTTGT GTATAGACAT AAAAGTGGGA AATTAAGGTA ACAAAGAGAA
TATGGGTTGA ATACAAAAAC ACACAGACAT ATATCTATTT GTAATGTATG GGTATAAACA CATATCTGTA TTTTTCACCT TTAATTCCAT TGTTTTCTCT

20 40 60 80 100
ATGGGAAGAG GAAGAGTAGA GCTGAAGAGG ATAGAGAACA AAATCAACAG ACAAGTAACG TTGCAAAAGC GTAGGAACGG TTTGTTGAAG AAAGCTTATG
TACCCTTCTC CTCTCATCT CGACTTCTCC TATCTCTGT TTTAGTTGTC TGTTCATTGC AAACGTTTCG CATCTTGCC AAACAACCTC TTTGGAATAC

120 140 160 180 200
AATTGTCTGT TCTCTGTGAT GCTGAAGTTG CTCTCATCAT CTCTCCAAC CGTGGAAAGC TCTATGAGTT TTGCAGCTCC TCAAGTAAA CAACTCTCTC
TTAACAGACA AGAGACACTA CGACTTCAAC GAGAGTAGTA GAAGAGGTTG GCACCTTTTCG AGATACTCAA AAGTCGAGG AGTTTCAATT GTTGAGAGAG

220 240 260 280 300
ACTCTTATC AGTTTCTTGA TTGAGTTTT GCTAGATCTG AGCTTAGATC TTTGCTCAA GGACTTGTTA TATATAGATC ACACGATCTT GATTCTTACG
TGAGAAATAG TCAAGAACT AACTCAAAAA CGATCTAGAC TCGAATCTAG AAACAGAGTT CTTGAACAAT ATATATCTAG TGTGCTAGAA CTAAAGATGC

320 340 360 380 400
AAGTTGAGTT AATTAGATTT CTTGATTCA TTTTCTAGGG TTTTTCCTCA ATCTTGAAA TTTAAGATCT GGTTTTTTGT TTGTCATGA TTTAGAACTG
TTCAACTCAA TTAATCTAAA GAACTAAAGT AAAAGATCCC AAAAAAGGT TAAGAACTTT AAATCTAGA CCAAAAAAC AACGTTTACT AAATCTTGAC

420 440 460 480 500
TGAAATTTGT AATCGAATAG ATTCCAAATC CTGATATGCA ATCTGAAAAG TTTTATATAA TTAATATATG TCTGTGTGAT TGGAACTTA AAAGTTGTTC
ACTTAAACA TTAGCTTATC TAAGGTTTAG GACTATACGT TAGACTTTTC AAAATATATT AATTATATAC AGACACACTA ACCTTTGAAT TTTCAACAAG

520 540 560 580 600
ACAGATTCT ATGAAAAATA CAAGTATCCA ACGTAGAATG ATAATATATG GTTACATGCA TTAACCAATT GTTAGTTCAT CATACTTTAT GGTGGTTAAA
TGTCTAAAGA TACTTTTAAAT GTTCATAGGT TGCATCTTAC TATTATATAC CAATGTACGT AATTGGTAAA CAATCAAGTA GTATGAAATA CCACCAATT

620 640 660 680 700
ACTTCAAACG CGTGATATC TGTGAAGGCT TTGATTGTTT GTTTTTCTT AAAAAAATG TTTAATAGAT TTTTAATTAT ATGTTAAAA AGTTTGTCTT
TGAAGTTTGC GCACATATAG ACACCTCCGA AACTAACAAA CAAAAAGAA TTTTGTGTAC AAATTATCTA AAAATTAATA TACAATTTTA TCAAAACGAA

720 740 760 780 800
ACATGCATTC AAGAAAATAT AGCGATTAAAT TCCTTTTTTC AAATCACAAT TTGTGAATCA AACGAAAACG TAAGATATTG CTTGCAAATG ATAGGATTGA
TGTACGTAAG TTCTTTTATA TCGCTAATTA AGGAAAAAAG TTTAGTGTTA AACACTTAGT TTGCTTTTGC ATCTATAAC GAACGTTTAC TATCCTAACT

820 840 860 880 900
ACTATTGATA TTGTAAATA TAAATACGAA ACTTTACGTT TGAAGTTGA AACAATCAAA TCCAAATCAA CTCGTATATA ATCAGATAAA TAATGGAAC
TGATAACTPAT AAACATTTAT ATTTATGCTT TGAATGCAA ACTTTCAACT TTGTTAGTTT AGGTTTAGTT GAGCATATAT TAGTCTATTT ATTACCTTTG

920 940 960 980 1000
AATCTTCAAT TTGTATGAA GAATACCTTA AAACCTGAAG AGCTTTTTTT TTATGGTGAT TTATAGGTTT AGATCTCCAA AGTCAAGTAT GATCTTTTTA
TTAGAAGTTA AAACACTCTT CTATGAAAT TTTGAACCTC TCGAAAAAAA AATACCACTA AATATCCAAA TCTAGAGGTT TCAGTTTATA CTAGAAAAAT

1020 1040 1060 1080 1100
ATAAACTCTT ATTCTCTCTT TTTGAGTTAT TTTCAGCATG CTCAAGACAC TTGATCGGTA CCAGAAATGC AGCTATGGAT CCATTGAAGT CAACAACAAA
TATTTGAGAA TAAGAGAGAA AAACCTCAATA AAGTCGTAC GAGTCTGTG AACTAGCCAT GGTCTTTACG TCGATACCTA GGTAACTTCA GTTGTGTGTT

1120 1140 1160 1180 1200
CCTGCCAAAG AACTTGAGGT GTTCTTAATT CAAATACTAT TTTAGATTCC TATCATATCA TTTCAAGAAA GATCTTTTTT AAAAGTTTGT TTTCTGTAAA
GGACGGTTTC TTGAACCTCA CAAGAAATTA GTTTATGATA AAATCTAAGG ATAGTATAGT AAAGTTCTTT CTAGAAAAAA TTTTCAACA AAAGCACTTT

1220 1240 1260 1280 1300

TATTTCAGAA CAGCTACAGA GAATATCTGA AGCTTAAGGG TAGATATGAG AACCTTCAAC GTCAACAGAG GTACATATCT GTCTACCTCC GTATATTIAC
ATAAAGTCTT GTCGATGTCT CTTATAGACT TCGAATTCCC ATCTATACTC TTGGAAGTTG CAGTTGTCTC CATGTATAGA CAGATGGAGG CATATAAATG

1320 1340 1360 1380 1400
TCAATTCTGT ATCCATGTAG ATTCATATTT GTAGGTGTGT GTGGCTTTTG TTGGTGAGAG AATCTTCTTG GGGAGGATTT AGGACCTTTG AATTCAAAGG
AGTTAAAGCA TAGGTACATC TAAGTATAAA CATCCACACA CACGAAAAAC AACACGCTCT TTAGAAGAAC CCTCTCTAAA TCCTGGAAAC TTAAGTTTCC

1420 1440 1460 1480 1500
AGTTAGAGCA GCTTGAGCGT CAACTGGAGC GCTCTCTCAA GCAAGTTCGG TCCATCAAGG TATCTTTATA CATGGAATCA ATGATTCAAA TGAGATTAAT
TCAATCTCGT CGAATCGCA GTTGACCTGC CGAGAGAGTT CGTTCAAGCC AGGTAGTTCC ATAGAAATAT GTACCTTAGT TACTAAGTTT ACTCTAATTA

1520 1540 1560 1580 1600
TTGTGTTGTT TAATTATAAC TACTATGGTG GTATGATGAT TGTMTGCAGA CACAGTACAT GCTTGACCAG CTCTCGGATC TTCAAAATAA AGAGCAAATG
AACACAACAA ATTAATATTG ATGATACCAC CATACTACTA ACAACGCTCT GTGTCAATGA CGAATGGTC GAGAGCCTAG AAGTTTATT TCTCGTTTAC

1620 1640 1660 1680 1700
TTGCTTGAAA CCAATAGAGC TTGGCAATG AAGGTATAAT TACAGAAATA ATGCATTTGG TGCCTTGCGA TCAATCTCTT TCACAGAGTT TAAGTTTCTA
AACGAACITT GGTATCTCG AAACCGTTAC TTCCATATTA ATGTCTTATT TACGTAAACC ACGGAACGCT AGTTAGAGAA AGTGTCTCAA ATTCAAAGAT

1720 1740 1760 1780 1800
AACATTTTGG GAAACATCTC TAGTTTTCTT GTTCTGATT ATAGTCTTTT GGTGAAATGT AAATGTTTAG CTGGATGATA TGATTGGTGT GAGAGTCAT
TTGTAAAAAC CTTGTAGAG ATCAAAAGAA CAAAGACTAA TATCAGAAAA CCACTTTACA TTACAAATC GACCTACTAT ACTAACCACA CTCTTCAGTA

1820 1840 1860 1880 1900
CATATGGGAG GAGGAGGAGG ATGGGAAGGT GGTGAACAGA ATGTTACCTA CGCGCATCAT CAAGCTCAGT CTGAGGAGT ATACCAGCCT CTTGAATGCA
GTATACCCCT CTCCTCTCC TACCCTTCCA CCACTTGCT TACAATGGAT GCCTGTAGTA GTTCGAGTCA GAGTCCCTGA TATGTCGGA GAACCTACGT

1920 1940 1960 1980 2000
ATCCAACTCT GCAATGGGG TAAATCCTTT GCCTTAAACA ATCATCTGCA AATCAGCTTG TGTACTTCAC TACTAAGATT GTACTTATAT AAGGTTCTTT
TAGGTTGAGA CGTTTACCCC ATTTAGGAAA CGGAATTTGT TAGTAGACGT TTAGTCGAAC ACATGAAGTG ATGATTCTAA CATGAATATA TTCCAAGAAA

2020 2040 2060 2080 2100
AGTTACTTGG TGTAAAGAGG ATCATCAATG TGTGTGAACC TTTTAAGTTG CTGTTTTGGT GATGATGATG ATGATGACAG GTATGATAAT CCGGTATGCT
TCAATGAACC ACATTTCTCC TAGTAGTTAC ACACACTTGG AAAATTCAAC GACAAAACCA CTACTACTAC TACTACTGTC CATACTATTA GGCCATACGA

2120 2140 2160
CAGAGCAAAT AACTGCGACA ACCCAAGCTC AGGCGCAGCA GGGAAACGGT TACATCCCGG GGTGGATGCT C
GTCTCGTTTA TTGACGCTGT TGGGTTGAG TCCGCGCTGT CCTTTTGCCA ATGTAGGGCC CCACCTACGA G

SEQ ID NO:42

Arabidopsis SEP2 genomic sequence

-2981 -2961 -2941 -2921 -2901
ACGCTCTAAC CAACTGAGCT AATGGGCCAT TTGCGAATGG TAGTGTCTAT TTTACTTATT CGAATCTAAA TCGTCTAGAG TAATTAAGAA GACATGCAAA
TGCAGATTG GTTGACTCGA TTACCCGGTA AACGCTTACC ATCAGAGATA AAATGAATAA GCTTAGATTG AGCAGTATCC ATTAATTCTT CTGTACGTTT

-2881 -2861 -2841 -2821 -2801
GCCTAATCAA TGATGGATTC TTGATTCTA CTCTAGGTG CCACCATGGA CGCATTGATA AAATCATAAC CGGTGCTTTA CAAAACATAT TGCTTGAATG
CGAATTAGTT ACTACCTAAG AAACCTAAGAT GAAGATCCAC GGTGTAACT GCGTAAGTAT TTTAGTATTG GCCAGCAAAAT GTTTTGATTA ACGAACTTAC

-2781 -2761 -2741 -2721 -2701
ATTCTAAACA AATAATAGTT TTTGTTGAA ATTTTCAAAA CATATGTTAG GTAAGGTGAG GTTTTGCCAA TAAGCCTTAC TATATACAGT GGCAACATGT
TAAGATTGTT TTATTATCAA AAAACAACCT TAAAGATTTT GTATACAATC CATTCCAGTC CAAAACGGTT ATTCGGAATG ATATATGTCA CCGTTGTACA

-2681 -2661 -2641 -2621 -2601
TTCTTCTACT TTGAGGAGTT TTGGGTGAAT ATGAAACCCA TGTGAGCATG ATACATGTGT TTCTTCTTCT ATTGAAATTT CCCCCAATGG TCATTTGCTC
AAGAAGATGA AACCTCTCTA AACCCACTTA TACTTTGGGT ACACCTCGTAC TATGTACACA AAGAAGAAGA TAACCTTTAA GGGGGTTACC AGTAAACGAG

-2581 -2561 -2541 -2521 -2501
TTTGCCTTCG TGTGCGCTT TCCGGTATCA AATCATATAT ATATATAACC TAAATGAGAG TAGACAATTT GAATCATTGT AAAAGGTATA AAGAAGAGAT
AAACCGAAGC ACAACGCGAA AGGCCATAGT TTAGTATATA TATATATTGG ATTTACTCTG ATCTGTAAAC CTTAGTAAAC TTTTCCATAT TTCTTCTCTA

-2481 -2461 -2441 -2421 -2401
TATAGTCCAC AATTAAACAA GTAATAAGAC GGTAAATAT CAAAACAAAT GAAAGGGTAA AAAAAAACA AGAGGGACAA GTCACCTGTTA GAAAGGTGAC
ATATCAGGTT TTAATTGTTT CATTATCTCT CCAATTTTATA GTTTGTTTAA CTTCCTCAT TTTTCTTGT TCTCCCTGTT CAGTGACAAAT CTTCCTACTG

-2381 -2361 -2341 -2321 -2301
TCCTCCCTTT GGGCCAGCCC CCTACCACAA AAGTCAAGGC TTACTTACTA TTCAGTCATA TATCGACAGG TGACTTCCA ACCACATCAC CCATCCTATT

AGGAGGGAAA CCCGGTCGGG GGATGGTGT TTCAGTTTCG AATGAATGAT AAGTCAGTAT ATAGCTGTGC ACATGAAGCT TGGTGTAGTG GGTAGGATAA

-2281 -2261 -2241 -2221 -2201

ACGTAATTC CACTGTCTAG ACTTTTTTTT TTTTTTTTTT TTTTACTTTT TAACGTTTTT TAGCTGTCTC TCTAAATTAC TACATACGGA CTTGCTACGT
TGCATTAAAG GTGACAGATC TGAAAAA AAAA AAAA AAAA ATTGCAAAAA ATCGACAGAG AGATTTAATG ATGTATGCCT GAACGATGCA

-2181 -2161 -2141 -2121 -2101

CACCTGAGAA GAAAGATCTT TGCTCGTAG TCTTTTGTCT GAAGAAAAAT TATTGTATTT TAGTTATTTA CAATTGCGATA ATTGTGTGTA GTAAATCCGC
GTGGACTCTT CTTTCTAGAA ACGAGCATCT AAGAAACAGA CTTCCTTTTA ATAAACATAA ATCAATAAAT GTTAACGTAT TAACACACAT CATTTAGCGG

-2081 -2061 -2041 -2021 -2001

CAGAATGATA TTAGAGTGAAT ACTGAGACGA CGAATGTGT AACTTGTAACT ATATATACTA ATAAACACGA TTGATTAAAA ATTACTATA CAGTATATCC
GTCTTACTAT AATCTCACTA TGACTCTGCT GCTTACCACA TTGAACATTG TATATATGAT TATTGTGTCT AACTAATTTT TAAATGATAT GTCATATAGG

-1981 -1961 -1941 -1921 -1901

AAAAATTAT GATTGAGAT GTACATATAC AATAAGTAAT TAAACCTCAA AACCAACAG TTTTTTTTTT TTTTGGTCAA CAATAATTAG AAATGAGAAT
TTTTGTAATA CTAACCTCA CATGTATATG TTATTCATTA ATTTGGAGTT TTGGTTTGTG AAAAAA AAAACCAAGT GTTATTAATC TTTACTCTTA

-1881 -1861 -1841 -1821 -1801

AAACTATTTA ACTTATAAAT TCTAGACCCA AAACTCATA TTTTACCCTT CTGTGTCTCA CTTAAAGA CTTAATTCC CAAAACCTCT GCAACAATAG
TTTGATAAAT TGAATATTTA AGATCTGGGT TTTTGAGTAT AAAATGGGAA GAACCAAGT GGATTTTCTT GAAATTAAGG GTTTTGAGAA CGTTTGTATC

-1781 -1761 -1741 -1721 -1701

GCCAAACATA GAAGATTGGA AAACAAATTT AAATCTACTT TCACTTTTAT AAAGAATAAT CAACGAACCA ATTAAGTTAA ACCTACATAT ATTGATATGT
CGGTTTGTAT CTTCTAACCT TTTGTTTAAA TTTAGATGAA AGTGAAAAATA TTTCTTATTA GTTGTCTGGT TAATTCATTT TGGATGTATA TAAGCATACA

-1681 -1661 -1641 -1621 -1601

GATCAGATAT GTGTTATATT CCTCAGCTT TCTTCCATTT AGCTAATAAC CTTAATTACT TCAAGAAATC ATATATCAAC CGAAAACCTAG TAAATAAAT
CTAGTGATA CACAATATAA GGAGTGCAAG AGAAGGTAAA TCGATTATTT GAATTAATGA AGTTCTTTAG TATATAGTTG GCTTTTGATC ATTTTATTTA

-1581 -1561 -1541 -1521 -1501

ATACATACTG AAAGCGGCA AAATTTTITAG CAATATTTTA AAATACCCTA CATCATAGTC TTAACATAAT AATCTTTCTG ATCAAAATTT ATTTTCATAA
TATGTATGAC TTTCCGCGCT TTTAAAAATC GTTATAAAT TTTATGGGAT GTAGTATCAG AATTGATTAA TTAGAAAGAC TAGTTTAAA TAAAGATATT

-1481 -1461 -1441 -1421 -1401

TATTCATAAA TACTTTAGGA TTACCTAAAC CAGGATACTT ATCCCTATAA ATCTGTCAAT CATCATGGAT TCATGGAGAC ATGGTCAGAT ATCCACGTC
ATAAGTATT ATGAATACCT AATGGATTG GTCTATGAA TAGGGATATT TAGACAGTTA GTAGTACCTA AGTACCTCTG TACCAGTCTA TAGGGTGACG

-1381 -1361 -1341 -1321 -1301

CAGATACAAT GTAACATAAT GATATACTGC GGCTGATTAT TATTTTITAC ATTAGAACGA GTTTAGATCC AAAACAAAT TGGTATTCTC AAACAAAAAT
GTCTATGTTA CATTTGATAA CTATATGACG CCGACTAATA ATAAAAATG TAATCTTGCT CAAATCTAGG TTTTGTTTTA ACCATAAGAG TTTGTTTTTA

-1281 -1261 -1241 -1221 -1201

TAAAAATTGA ATACGAAAGT AATAGAACAA AACTTCAATG TTGTGGAATA GATAGGAAGC AATAGAAAAG CGACACGTAC ATGTCCATTT TAAGGTAGGA
ATTTTAACT TATGCTTTCA TTATCTTGT TTGAAGTTAC AACAGCTTAT CTATCCTTCG TTATCTTTTC GCTGTGATG TACAGGTAAA ATTCACCTCT

-1181 -1161 -1141 -1121 -1101

GAGGCTTTTC TCGGCTTGT GAAGTAAGAA AAAGAAAATG ATGATAGCTG CTTTCTTTTC ATTCATTGCA GAAGAAACCA ATGTTTCCCC AATCTCACGC
CTCCGAAAAG ACGCGAACA CTCATTCTT TTTCTTTTAC TACTATCGAC GAAAGCAAAG TAAGTAACGT CTTCTTTGTG TACAAAGGGG TTAGAGTGGG

-1081 -1061 -1041 -1021 -1001

GCCTCCTCCT ATCTACCACC ACTTGGACAA ATCCCTTTT CAGTATTAGT TTTTITTTCC GGACATTGTA CATTCAAAAG CATTCCAAGT GTCTAATAAA
CGGAGGAGGA TAGATGGTGG TGAACCTGTT TAGGGGAAAA GTCATAATCA AAAAAAAGG CCTGTACAT GTAAGTTTTT GTAAGGTCTA CAGATTATTT

-981 -961 -941 -921 -901

CATAACTAAC CACTCCAAGA TGCAAAATCT AGCTACGAAC AAATTTTAAA CTATAGAGAT GAACCTTAAA TTCCGGCAAT AATTAGTGA ACTTGAGCTA
GTATTGATTG GTGAGGTCTC ACGTTTTAGA TCGATGCTTG TTTAAATTT GATATCTCTA CTTGAAATTT AAGCCCGTAA TTAATCACCT TGAACCTGAT

-881 -861 -841 -821 -801

TTGATGAGTT TTCTGACTTT TTGAAGCTTA ATTGAGTTTT ATATACACTA TATATAGGCT TGAATAATA TGGATCAAC AAGAAATATA TAACTACAA
AACTACTCAA AAGACTGAAA AACTTCGAAT TAACTCAAAA TATATGTGAT ATATATCCGA ACATTATTAT ACCTAGTTTG TCTTTTATAT ATTTGATGTT

-781 -761 -741 -721 -701

AITGGGAATT AGGTTTTTAA ACGTTATCGT TCTATTTTAA TTCAGGCACC TTTAGAATAT CAAGATCCAT GCATGTTTCA ATATTCTGT TGACAAATAA
TAACCTTAA TCCAAATTT TGCAATAGCA AGATAAAAT AAGTCCGTGG AATCTTATA GTTCTAGSTA CGTACAAAG TATAAGACA ACTGTTTATT

-681 -661 -641 -621 -601

ATAAAGATGT CTCAAATATG AAGTTTGGGC AAGTACGTG TAGACCTAAA AGAGTCGAAA CATTGGTATC TAAGTCATAT ATCTAGATGT ATATGACAT

5

10

15

20

25

30

35

40

45

50

55

60

65

TATTTCTACA GAGTTTATAC TTCAAACCCG TTGCATGCAC ATCTGGATT TCTCAGCTTT GTAACCATAG ATTCACTATA TAGATCTACA TATACCTGTA
 -581 -561 -541 -521 -501
 GGATTATATA ACTAGACAA GTTTGTTTA AAACTTAAT TCATTTTCT TAATTAGTAG CAACTAGCAA CTAACCTACT ATGGCAAATA ATGGTGTCTG
 CCTAATATAT TGATCTGTG CAAACAAAT TTTTGAATTA AGTAAAGA ATTAATCATC GTTGATCGTT GATTGATGAG TACCCTTIAT TACCACAGAC
 -481 -461 -441 -421 -401
 CGTGGCAGC ACTTGGAGA GAAGGTGTGA GAATGTTTT TACTTTCTGT GTAAAGATG GAAGAGAGAG AAAGAGTAA GAAGTAGAGA GAGAGATATT
 GCACCGTGC TGAACCTCT CTTCACACT CTTACAAAA ATGAAGACA CATTTCTAC CTTCTCTCTC TTCTCATTT CTTCATCTCT CTCTCTATAA
 -381 -361 -341 -321 -301
 GTATCACAA ACCCTAATGA TCTCTACCC TCACAAATTT TCTTATCTTT ATAGCTTTTA TAGATTACA AAACTTTTC TTCAGATTCA CAATCTCATC
 CATAGTGGTT TGGGATTACT AGACAGTGG AGTGTTTAAA AGAATAGAAA TATCGAAAAT ATCTAAGTGT TTTTGAAAAG AAGTCTAAGT GTTAGAGTAG
 -281 -261 -241 -221 -201
 ACAACCTTC AAAAAGAGA AAGATCTAAA GAATAACAA GAGCCCTAAT ATCAAATCAC AACCAAAAA ACCAAAGAAA GCTAATTAAA GTTTTCTCTC
 TGTGGGAAG TTTTCTCTT TTTCTAGATT TTTATTGTT CTCTGGATTA TAGTTTAGTG TTGGTTTTTT TGGTTTCTT CGATTAATTT CAAAAGAGAG
 -181 -161 -141 -121 -101
 TAGCTATTCC TCCTCTTTT TTTTCTTGA AAAGTAGGTT TTACTTCACC AAAAGATAAG ATCTTTCCCC AGAAAAGCA ATACCCAAGT CATGTTCTCTG
 ATCGATAAGG AGAAGAAAAG AACAAGAACT TTTGATCCCA AATGAAGTGG TTTTCTATTC TAGAAAGGGG TCTTTTCTGT TATGGGTTC GACAAAAGAC
 -81 -61 -41 -21 -1
 TGTGTCTGTA TATAGATAAA ACATTACATA CCCTAATAGA GTTACACAAA TAGCTATAAA AGAGGGA AAAAGTAGGG ATTTTGGG GTGAGGAAA
 ACACAGACAT ATATCTATTT TGTAAATGAT GGGATTATTC CAATGTGTTT ATCGATATTT TCTCCCTTTT ATTCTATCCC TAAAAACCC CACTCTTTTC
 20 40 60 80 100
 ATGGGAAGAG GAAGAGTAGA GCTCAAGAGG ATAGAGAACA AAATCAACAG ACAAGTAGCG TTTGCTAAAC GTAGAAATGG TTTGCTGAAA AAAGCTTATG
 TACCCTTCTC CTCTCATCT CGAGTTCTCC TATCTCTGT TTTAGTTGTC GTTCACTGC AAACGATTGG CATCTTTACC AAACGACTTT TTTGGAATAC
 120 140 160 180 200
 AGCTTTCTGT TCTCTGCGAT GCTGAAGTCT CTCTCATCGT CTCTCCAC CGTGGCAAGC TCTACGAGTT CTGACAGACC TCCAAGTACT TCTCTTCTT
 TCGAAAGACA AGAGACGCTA CGACTCAGA GAGATAGCA GAAGAGGTTG GCACCGTTCG AGATGCTCAA GACGTCTGG AGGTTCTATGA AGAGAAAGAA
 220 240 260 280 300
 TATACACTTA TTAGATCTGT GTGTAGATCT TTTATTTTC TAGTCTTGTG ATGAGTTTGA TCTTTCTTGA TTGCTTTTGA ACAAATACT TGATATATTT
 ATATGTGAAT AATCTAGACA CACATCTAGA AAGTAAAGAG ATCAGAACAC TACTCAAAAT AGAAAGAACT AACGAAAAT TGTTTTATGA ACTATATAAA
 320 340 360 380 400
 TCAGTTTCTT AATCTGATCT CTAATTAGGT TTTGATTATA GAAGAATAAT TCAGTACTTT CAAGTGATTG AATTTTCGAGA TCTGATCTTA ATTTATCAT
 AGTCAAGAAA TTAGACTAGA GATTAAATCCA AAATAATAT TTTCTTATTA AGTCATGAAA GTTCACTAAC TTAAGGCTCT AGACTAGAAT TAAATTAGTA
 420 440 460 480 500
 CATGTCAAAT TCTTAGGGAT TTAATTGCAA TCTATTTTGA GATTATCGG AGCTAGGAAA STATCATAAT GATATACTAT TATTATCATG TAAITTCATT
 GTACAGTTTA AGAATCCCTA AATTAACTGT AGATAAAAT CTAATAGCC TCGATCTTTT CATAGTATTA CTATATGATA ATAATAGTAC ATTAAAGTAA
 520 540 560 580 600
 GTCTCTACAC GGATATATAT GTGATTAGAA TTTGGTAAAG TAACTAAAG ATTCACAGTC TTCAATGAAA TTTAAAGAT CCAACGTAGA ATAATTAGTG
 CAGAGATGTG CCTATATATA CACTAATCTT GAACCATTTT ATTTGATTTT TAAGTGTGAG AAGTTACTTT AAATTTTCTA GGTGCGATCT TATTAATCAC
 620 640 660 680 700
 GTTCCATGCA TTAACCACTC TAAATAAGC TCATGCAGAC ATTTAAGCAC CACATGAATT TAATATCTTT TTAATTAAAG GATCTTCTTT TTATAAATTT
 CAAGGTACGT AATTGGTCAG ATTAATTTCTG AGTACGCTGT TAAATTCGTG GTGTACTTAA ATTATAGAAA AATTAAITTC CTAGAAGAAA AATATTAAAA
 720 740 760 780 800
 TCTTTTGTGA GTTTTAAAA TTTTAGTTTG TTTATTAAT TTATAGATTC TTCTCTCTCT GATTGTGTTT TTTTGTCTT TCAGCATGCT CAAGACACTG
 AGAAAAAAT CAAAAATTT AAAATCAAA AAGTAATTTA AATATCTAAG AAGAAGAGGA CTAAACACAA AAACTAGAAA AGTCGTACGA GTTCTGTGAC
 820 840 860 880 900
 GAAAGGTATC AGAAGGTGAG CTATGGCTCC ATTTGAAGTCA ACAACAAACC TGCTAAAGAG CTGTAGGTTT AATCTCCAAC ATCTCTTCGA TCTTAATTAAT
 CTCTCCATAG TCTTCACATC GATACCGAGG TAACTTCAGT TGTGTTTGG ACGATTCTCT GAACTCCAAA TTAGAGGTTG TAGAGAAGCT AGAATTAATA
 920 940 960 980 1000
 TTATCTTTT TTAATTTTAT CTAAGAAAA TGTGTTGATT TGAGACAAAA GCCCTTCAAA GTTTCTTACA TAGATATTCA ATTGTCTATT ATCTTCGCAA
 AATAGGAAAA AATTAATAAT GATTTCCTTT ACAAACTAAA ACTCTGTTT CCGGAAGTTT CAAAGAATGT ATCTATAAGT TAACAGATAA TAGAAGCGTT
 1020 1040 1060 1080 1100
 TTTTCAGAAC AGTACTAGAG AGTACTTGAA GCTGAAGGAT AGATATGAAA ATCTGCAACG TCAGCAGAGG TATATACATT AATGTGGATG ATGATCATTT

AAAAGTCTTG TOGATGTCTC TCATGAACTT CGACTTTCCA TCTATACTTT TAGACGTTGC AGTCGTCTCC ATATATGTAA TTACACCTAC TACTAGTAAA
 1120 1140 1160 1180 1200
 ATAAACAGCA TATATATATA TATATATATA TATATATATA GTTTGTATTG ATCATGAAAG TGTGTTGCTG CAGAAATCTT CTGAGAGAGG ATCTTGGACC
 TATTTGTGCT ATATATATAT ATATATATAT ATATATATAT CAAACATAAC TAGTACTTTC ACACAACGAC GTCTTTAGAA GAACCTCTCC TAGAACCTGG
 1220 1240 1260 1280 1300
 TCTGAATTCA AAGGAGCTAG AGCAGCTTGA GCGTCAACTA GACGGCTCTC TGAAGCAAGT TCGCTGCATC AAGGTGATTT ACITCTGTAC ATACACTGAA
 AGACTTAAGT TTCTCGATC TCGTCGAACT CGCAGTTGAT CTGCCGAGAG ACTTCGTTCA AGCGACGTAG TTCCACTAAA TGAAGACATG TATGTGACTT
 1320 1340 1360 1380 1400
 AGATTACAC AAATCTTTCT CTATATATAG ACTGAGACAC ATGCATGAAA TGTTTGTGAT GCGTGAGGTT ATCTGAAATG GGCTCTTCTT TTTTGGAGCC
 TCTAAGTGTG TTTAGAAAGA GATATATATC TGACTCTGTG TACGTACTTT ACAAAAACCTA CGCACTCCAA TAGACTTTTA CGGAGAAGAA AAAACGTCTG
 1420 1440 1460 1480 1500
 ACAGTATATG CTGAGCAGC TCTCTGATCT TCAAGGTAGG GAGCATATCT TGCTTGATGC CAACAGAGCT TTGTCAATGA AGGTATATGA TGTGTTTCT
 TGTATATATC GAACTGGTGC AGAGACTAGA AGTTCCATTC CTCGTATAGA ACGAACTACG GTTGTCTCGA AACAGTTACT TCCATATACT ACTACAAAGA
 1520 1540 1560 1580 1600
 CTCTCTCTCC TCCAGTTTCT ATTTATAGAT GGAACTTTTA AATAGTCCAA TTTATATATA TGAGTCTAAA TTTCACATTC TTCAACTGCT ACATGTTTCT
 GAGAGAGAGG AGGTCAAAGA TAAATATCTA CCTTTGAAAT TTATCAGGTT AAATATATAT ACTCAGATTT AAAGTGTAAG AAGTTGACGA TGTACAAAGA
 1620 1640 1660 1680 1700
 TTTGTATTAT TTCTATGATA TCTTCAGGAA AGTTTGAAAA ATATTGTGTT TTGTTTAGCT GGAAGATATG ATCGGGGTGA GACATCACCA TATAGGAGGA
 AAACATAATA AAGATACTAT AGAAGTCTCT TCAAACTTTT TATAACACAA AACAAATCGA CCTTCTATAC TAGCCGCACT CTGTAGTGGT ATATCTCTCT
 1720 1740 1760 1780 1800
 GGAATGGAAG GTGGTGATCA ACAGAAATAT GCCTATGGAC ATCCTCAGGC TCATCTCTAG GCACTATACC AATCTCTTGA ATGTGATCCC ACTTTGCAAA
 CCTACCCTTC CACCACTAGT TGTCTTATAA CGGATACCTG TAGGAGTCCG AGTAAGAGTC CCTGATATGG TTAGAGAACT TACACTAGGG TGAACGTTT
 1820 1840 1860 1880 1900
 TTGGGTAAAT CAAACAACCT TTCTTCCTT AAGACATCAA CTTAGGTTAT AAACAGTTAG CAGTTTGCTT TAAGCCCAAC ATTTGCTTTG TTTCTAGAG
 AACCCTTTTA GTTTGTGTAA AAGAACGGAA TTCTGTAGTT GAATCCAATA TTTGTCAATC GTCAAACGAA ATTCGGGTG TAACAGAAAC AAATATCTC
 1920 1940 1960 1980 2000
 GCTTTGGTTA AAATCGTGT TGTTTAGTCT AAGGATTCAG CACTTTGATG TCTGAAGTAT GGAAATCAA TATCTCAGAC TTGAAATGT GGGTTTCTAT
 CGAAACCAAT TTTGAGCACA ACAAATCAGA TTCTTAAGTC GTGAAACTAC AGACTTCATA CCTTTTAGTT ATAGAGTCTG AACTTTTACA CCCAAAGATA
 2020 2040 2060 2080 2100
 TGTGTACTTC GAAACTATGT TGTGTGGTGT TTGCAACAG ATATAGCCAT CCAAGTGTCT CAGAGCAAAAT GGCTGTGACG GTGCAAGGTC AGTCCCAACA
 ACAACTGAAG CTTTGATACA ACAACCCAC AACGTTGTCT TATATCGGTA GGTACACAGA GTCTCGTTTA CCGACACTGC CACGTTCCAG TCAGGGTGTG
 2120
 AGGAACCGCG TACATCCCTG GCTGGATGCT G
 TCCTTTGCCG ATGTAGGGAC CGACCTACGA C

SEQ ID NO:43

Arabidopsis SEP3 genomic sequence

-2981 -2961 -2941 -2921 -2901
 GTCCCTTCC CATACGCTC TGACGTGGAC CCGTCCGTC TATTTTAGC AGATTAAATC AACGGTCTT ATTCTTTCTT CGACCTTCA CGACATTCGC
 CAGGGGAAGG GTAATGCAGA ACTGCACCTG GGACAGGCAG ATAAAAATCG TCTAATTAGG TTGCCAAGAA TAAGAAAGAA GCTGGGAAGT GCTGTAAACGG
 -2881 -2861 -2841 -2821 -2801
 TCAAAGCGGT CGATTCTCA TCTCAGGCC AATGACCAC ATATATCACC AGTACTCGC AACTTAGCTG TCGTGTAGGA TTTCAGTGG CATTTATTG
 AGTTTCGGCA GGCTAAGAGT AGAGTGGGG TTACTGTGTG TATATAGTGG TCATGAGCGG TTGAATCGAC AGCACAATCT AAAGTGCACC GTAATAAAC
 -2781 -2761 -2741 -2721 -2701
 TTCTAGTTTG TAGTGCAAC ATTGCAAGTT GATATGTGC CCTATCGATC ACCGTCGTCT CTTTAGCTTC ACATCGAGAT TCTTCTTTCT TTCTACGTG
 AAGATCAAAC ATCAGCTTTG TAACGTTCAA CTATACCAGG GGAATAGTAG TGGCAGCAGA GAAATCGAAG TGTAGTCTTA AGAAGAAAGA AAGGATGCAC
 -2681 -2661 -2641 -2621 -2601
 TAATAGCATT TTTGATTTTG AGAATTTCTT TAGAACCCTT GGATCTCTCA TCGTTGGTGG ATCCATCCAT CCAAATGGGA CTTGTGTGTG CTCCATCCAG
 ATTATCGTAA AAATCAAAAC TCTTAAAGAA ATCTTGGCAA CCTAGAGAGT AGCAACCAAC TAGGTAGGTA GGTTTACCTT GGACACACAC GAGGTAGTCT
 -2581 -2561 -2541 -2521 -2501
 GGCATATGAT CCCAAAGCCA AAAGAGTATT TCCAAGTCTT TTTCTTTCTT CTTTCTTTCT TTCTTACTAA CCTTTTCTTT TCTTATGCTT TAGACTAAGA
 CCGTATACTA CGGTTTCGGT TTTCTCTATA AGGTTACAGA AAGAAAGAAA GAAAGAAAGA AAGAAATGAT GGAAAAAAGA AGAATACGAA ATCTGATTTCT

5
10
15
20
25
30
35
40
45
50
55
60
65

```
-2481      -2461      -2441      -2421      -2401
AATTIATTCG GCCATATCCA CTTTACGAA TATACTTCTT ACAAGATCTA GATTITTTTG AGTTAATTCG GTGTATATAA CATTGGCATG GACTGCAATT
TTAAATAAGC CGGTATAGGT GAAAAATGCTT ATATGAAGAA TGTTCTAGAT CTAAAAAAC TCAATTAAAGC CACATATATT GTAACCGTAC CTGACGTTAA

-2381      -2361      -2341      -2321      -2301
AAGTAATGGT AATGTGATCA TGATGCGATG TGTCGTTATC AGTAGTATAA TATTGATGGG CTACCCTGGA AAACAAAATT ACGTGTATATA TGTACACAAT
TTCAATTACCA TTACACTAGT ACTACGCTAC ACAGCAATAG TCATCATATT ATAACTACCC GATGGGACCT TTTGCTTTTAA TGCACAATAT ACATGTGTTA

-2281      -2261      -2241      -2221      -2201
TTGGTAGAAC CGTAGAAATT AAACCTGAATA AAACCTTCTA TAATGTTCAA AATTATATGG TACAGATTAA TACGGAAAAA CATTACGCTT TTACGTAACA
AACCATCTTG GCATCTTTAA TTTGACTTAT TTTGGAAGAT ATTACAAGTT TTAATATACC ATGTTCTAATT ATGCTCTTTT GTAAAGTGCGA AATGCATTGT

-2181      -2161      -2141      -2121      -2101
ATTAAGTGGG AAGTAAAAAT ATCCCAAAAA TAATTATATC ACATCATTTG TATATTTCTA AGTTTTTTTA TATCTCTAAT GGTATATGTT TTACGAGATTG
TAATTCACCT TTCAITTTAA TAGGGTTTTT ATAAATATAG TGTAAGTAAAC ATATAAAGAT TCAAAAAAAT ATAGAGATTA CCATATACAA AATGTCTAAC

-2081      -2061      -2041      -2021      -2001
TTTTTTGGGA AAATCTTTAA AGAGACTTGA AGAATGTTTT TTTTATTTTT TCITGAAATG TTTGACACTT GAAACCGTTT AAAAECTCAA ATATAGTATA
AAAAAACCCCT TTTAAGAATT TCTCTGAACT TCTTACAAAA AAAAATAAAA AGAACTTTAC AAACGTGAA CTITGGCAAA TTTTGTAGTT TATATCATAT

-1981      -1961      -1941      -1921      -1901
TATCATTTGT GGTCTCATAC CTTGTAATTC ACCACATATA TTATCAATGG GGAAGATTTG AAAATTTTTG GGGGATCACA AAACGAAGGA AAGAGTACAA
ATAGTAACAA CCAGAGTATG GAACATTAAG TGGTGTATAT AATAGTTACC CCTTCTAAAC TTTTAAAAAC CCCCTAGTGT TTTGCTTCCT TTCTCATGTT

-1881      -1861      -1841      -1821      -1801
AAAGAGAAGG AAAAGATAGA AGATATATGT TTTTAACTTC ATTGGTATGA CATCAATAAA TAAATAGTTG AATGTACTTT AGTTTCTCTT TTGGTTTAAAT
TTTCTCTTCC TTTTCTATCT TCTATATACA AAAATTGAAG TAACCATACT GTAGTTATTT ATTATCAAC TTACATGAAA TCAAGAGAAA AACCATAATTA

-1781      -1761      -1741      -1721      -1701
GCACATCATC TCGATCAATT GTCATCATCT TACATTGAAT TATACGACCA GATCTGTATA CAAGTGAATT CGTACTTGCC CTTCCTTTTC TTCTCATACG
CGTGTAAGTAG AGCTAGTTAA CAGTAGTAGA ATGTAACTTA ATATGCTGGT CTAGACTATT GTTCACCTTAA GCATGAACGG GAAGGGAAAA AAGAGTAGTC

-1681      -1661      -1641      -1621      -1601
TCCTTCTAAC TAATTTTGAT TGTAACCTAT AATTATATAA CCATATTTAA TTTTATTTTA TCTAAAACCA ATTGAAGCAA ATTAAAAAT CATAAATCTT
AGGAAGATTG ATTAAAACTA ACATTGAATA TTAATATATT GGTATAAATT AAAATAAAAT AGATTTTGGT TAACCTCGTT TAATTTTATA GTATTTAGAA

-1581      -1561      -1541      -1521      -1501
GAGTCCACA TGAAGACAAT ATATAAACT CGTGCAAAAT TGCTTAAAT GCTTCTATGA GACCATGACC AAGTGAGATT AATAAGOGAT TCAATGTGCA
CTCAGGGTGT ACTTCTGTTA TATATTTTGA GCACGTTTAA ACGAATTTTA CGAAGATACT CTGGTACTGG TTCACTCTAA TTATTCGCTA AGTTACACGT

-1481      -1461      -1441      -1421      -1401
AATCAAAAGA GAAAGAAGC TAATGGGTTT AAATATAACC AAACAGAATA ATAATGCTAT GTTTAGTTTT TCTAATTGAA TCATACCTTT GTGTCCATCA
TTAGTTTCTT CTTTCTTTCG ATTACCCAAA TTTATATTGG TTTTCTTTAT TATTACGATA CAAATCAAAA AGATTAACTT AGTATGGAAA CACAGGTAGT

-1381      -1361      -1341      -1321      -1301
CCTACTTACC GGTGAGAATA AAGCAATTAC GTCTGCAACC AAAAAGCACT AAGACTTTTG GTGACACATG ATCTCTAACA TCGGACGAA CCTAAGATAA
GGATGAATGG CCACTCTTAT TTTGTTAATG CAGACGTTGG TTTTCTGTGA TTCTGAAAGC CAGTCTGTAC TAGAGATTGT AGCCTGCTTG GATTCTTATT

-1281      -1261      -1241      -1221      -1201
CCAAAAATAA CTATATCTTA TATTCAAATC TCTGTTTATT TTATCCATTT ATGTTTCTTT TCTTTCCCAT AATTTTTTTT GTGTCTCATC AGACTCTCTT
GGTTTTATTT GATATAGAAAT ATAAGTTTAG AGACAAATAA AATAGGTAAA TACAAAAGAA AGAAAGGGTA TTAAAAAAAA CACAGAGTAG TCTGAGAGAA

-1181      -1161      -1141      -1121      -1101
ACCAAACTGA ATTTATCAAC ATGGTTTTTT TTTTGGCCAC ATCAAAATGG TGGTTTATAA AGTAGACTAA TACAAAAGAC ATTTCTGTTA ATTTCACTAA
TGGTTTGAAT TAAATAGTTG TACCAAAAAA AAAACCGGTG TAGTTTACC ACCAAATATT TCATCTGATT ATGTTTTCTG TAAAGACAAT TAAAGTAGTT

-1081      -1061      -1041      -1021      -1001
CAAAAAAAT CTTAGCAGTA CTATAGATTG GAAAAGGAAA AGCAAACTCA GCAGTAAGAT TTATCAAAAC TAGCAGTAAG AGTTTATGAT ATCATGAAAA
GTTTTTATTA GAATCGTCAT GATATCTAAC CTTTCTCTTT TCGTTTAGAT CGTCATCTA AATAGTTTGG ATGCTCATTC TCAAAATCTA TAGTACTTTT

-981      -961      -941      -921      -901
CATCACAAAC GAGTAGTGT TTACTTTTACA TTTTAAACCA ATCACAAGGG TAGTCCGTA AGTTGGGAAA ATCGTACGAG GCTTCACCTA GTTAAAGTTA
GTAGTGTTTG CTCTACACAA AATGAAATGT AAAAATGGT TAGTGTCCC ATCAAGGCAT TCAACCTTT TAGCATGCTC CGAAGTGGAT CAATTCCAAT

-881      -861      -841      -821      -801
GGTCACATGA TTCCCTGAAC TCGATTTTAT AAGTAAAAA GAAAAATTTA TAAATCAAA ATTTTATATA TAAAAAATC AGGTGGATTT ATCAGACCTT
CCAGTGACTA AAGGAGCTTG AGCTAAAAATA TTCATTTTTT CTTTTTAAAT ATTTTAGTTT TAAAAAATAT ATTTTATTAG TCCACCTAAA TAGTCTGGGA
```


5
10
15
20
25
30
35
40
45
50
55
60
65

-781 -761 -741 -721 -701
 ACCATCGAGA TGTGCACACG TGTCCAACT CATTCAATGC CCTACTATTT TCTGTTTAGG GTTGCAATCA CTCATCGCAC ACGCGCCATC TCCACCTTCC
 TGGTAGCTCT ACAGCTGTGC ACAGGTTTGA GTAAGTAACG GGATGATAAA AGACAAATCC CAACGTTAGT GAGTAGCGTG TCGCGCGTAG AGGTGGAAGG
 -681 -661 -641 -621 -601
 ATTATTAATC TCTCATTTC AACATCACAC TCTTACGAAT CATACGATTT TAATATCTCT GTCTCTCTCA ACGTATTAAA TAAAAATGGT TTAAATGTT
 TAATAATTAG AGAGTAAAAA TTGTAGTGTG AGAATGCTTA GTATGCTAAA ATTATAGAGA CAGAGAGAGT TGCATAATTT ATTTTACCAC AAATTTACAA
 -581 -561 -541 -521 -501
 AGGGTTTTTT GTAGGATTTT CAATTATTAA TCTCTATAAT TCGATGAAC T AAGTAAAAA GCATCAAAC TTTCTGGCAG ATCACAATTT TCTCTAACT
 TCCCAAAAAA CATCCTAAAA GTTAATAATT AGAGATATTA AGCTACTTGA TTCAATTTTT CGTAGTTTGA AAGAACCGTC TAGTGTAATA AGAGATTGTA
 -481 -461 -441 -421 -401
 AAATATGGAC TGAATTTGAA AAATTAAACC ACTAGCTAGA ATAAAGTGTG GGTGAGAGTG GAATCTAAT TTCTCTCTTT TACTAATTAT GTATAAACAC
 TTTATACCTG ACTTTAACTT TTTAATTTGG TGATCGATCT TATTTACAA CCACTCTCAC CTGAGATTA AAGAGAGGAA ATGATTATA CATATTTGTG
 -381 -361 -341 -321 -301
 AAAAAATGCAC CAAATTTTTA GGTGTAATAA TATCTAAGCA TGGTAGGGT AATTAACATT TTTCTTTTCA ATTTTGAAT ATTTGAATA ATCCTATGAG
 TTTTACGTG GTTTAAAAAT CCAAACTTTT ATAGATTTCG ACCTATCCCA TTAATTGTAA AAAAGAAAGT TAAACGTTA TAACTTATT TAGGATACTC
 -281 -261 -241 -221 -201
 GGTCTTTGGT ACACAATAAT TGGAGGATAT ATAGTTGAGT CTGAGAGTAT ATTAGAAAGA GAATATTCCA AGTAATGAAG CTGACATGTT TATATGTACT
 CCAGAAACCA TGTGTTATTA ACCTCCCATAT TATCAACTCA GACTCTCATA TAATCTTTCT CTATATAAGT TCAITACTTC GACTGTACAA ATATACATGA
 -181 -161 -141 -121 -101
 TTGAGAGAAG TGTGTGAGA TTGTACAAA TGTATATGTA CACTTTAAAA AGCAATATAA GATAGATAAA AAAAATATAA AGAAAAAAG AAAGAAAGAA
 AACTCTCTTC ACAACACTCT AAACATGTTT ACATATACAT GTGAAATTTT TCGTTATATT CTATCTAATT TTTTATATT TCTTTTTTCT TTCTTTCTCT
 -81 -61 -41 -21 -1
 AGAAAAAAG AGAGAGGCTC ATATATATAT AGAATTGCTT GCAAGGAAAG AGAGAGAGAG AGATTGAGAT ATCTTTTGGG AGAGGAGAAA GAAAAAGAAA
 TCTTTCTTTC TCTCTCCGAG TATATATATA TCTTAACGAA CGTCTCTTTC TCTCTCTCTC TCTAATCTA TAGAAAAACC TCTCTCTTTT CTPTTCTTTT
 20 40 60 80 100
 ATGGGAAGAG GGAGAGTAGA ATTGAAGAGG ATAGAGAACA AGATCAATAG GCAAGTGACG TTTGCAAAGA GAAGGAATGG TCTTTTGAAG AAAGCATACG
 TACCCTTCTC CCTCTCATCT TAATCTCTCC TATCTCTTGT TCTAGTATAT CGTTCAGTGC AAACGTTTCT CTCTCTTACC AGAAAAATTC TTTCGTATGC
 120 140 160 180 200
 AGCTTTTCACT TCTATGTGAT GCAGAAGTTG CTCTCATCAT CTCTCAAAT AGAGGAAAGC TGTACGAGTT TTGCAGTAGT TCGAGGTATA TATCTACTTT
 TCGAAAGTCA AGATACACTA CGTCTTCAAC GAGAGTAGTA GAAGAGTTTA TCTCTTTTCG ACATGCTCAA AAGCTCATCA AGCTCCATAT ATAGATGAAA
 220 240 260 280 300
 TGTATATATA TTAATTAATA CATAAACATT TTATATACAT ATTAAGTAAC ACAAATAATG CTGTATGTA TGGGCTCTCT TGTGATGTGT TGTGTGTCTG
 ACATATATAT AATGAATATT GTATTGTGTA AATATATGTA TAATTCATTG TGTTTTACA GAACATACAT ACCCAGAGAG ACATACACA ACAACACAGC
 320 340 360 380 400
 TACGTACGTG TTCTATCATA TCTTTTAAA AGAAGCAAAAG AGGAAAAAAA ATTTGGGATA CCCCATAATC GTATCATTTT ATAACAAGTT TGTCTTTTGT
 ATGCATGCAC AAGATAGTAT AGGAAAAATT TCTCTGTTTC TCTTTTTTTT TAAACCCATAT GGGGTTTAGA CATAGTAAAA TATTGTTCAA ACGAAAAAAC
 420 440 460 480 500
 ATGTTCTTTT GTGTTCTCTC TTGAATTCOA TTTTGTGTTT TGATTTTTTT TCTATTTCTC TTTACATCTA TCAAAGTTTT TTTTCTTATA TTTTATGTCT
 TACAAGAAAA CACAAAGAGA AACTAAAGGT AAAACAAAAA ACTAAAAAAA AGATAAGAG AAATGTAGAT AGTTTCAAAA AAAAGAATAT AAAATAACGA
 520 540 560 580 600
 TATTGTGTTG TCTACTTAAT TCACATTATC TGAGAGAAGA ACAATCTATC TGATATGAAA TTAGGGTTAA TTTCTCTTGT GAGTACTCTT TAATTCACAT
 ATAAACAAAC AGATGAATTA AGTGAATAG ACTCTCTTCT TGTTAGATAG ACTATACTTT AATCCCAATT AAAGAGAACA CTCATGAGAA ATTAAGTGTA
 620 640 660 680 700
 AAGCTTAAAG TTTCCACCTT TTGAATCTGG GGGTCTGCCA ATTCGATCAA ATCACTCAAT TTTGTGTGTA GATTGATATA AGTTTATAGG GGGATATTGT
 TTGCAATTTT AAAGTGGAA AACTAAGACC CCCAGCAGGT TAAGCTAGTT TAGTGAGTTA AAACAACAGT CTAATCTAT TCAAGTATCC CCCTATAACA
 720 740 760 780 800
 TTCCACGACA ATCCATTTTA GTAACCTTAA GGGGTTTCCA ATTTTGGGTT TTGAATTGAC GCTAATGTCA AATTCATCTA AAGTCCGTTG GATATGTATA
 AAGGTGCTGT TAGGTAAAT CATTTGGAAT CCCCAAGGTT TAAACCCCAA AACTTAACTG CGATTACAGT TTAAGTAGAT TTCAGGCAAC CTATACATAT
 820 840 860 880 900
 CTGCGGGATG GGATTCATCC TTTTCTCTGG GTTCTTTAGA TCTTCTCTTA AAAGACTAAC AGATTTTGTG GTAAACCCTA GGAACAGTT AAAAAATCCCA
 GAACCCCTAC CTAAGTAGG AAAAAGAGCC CAAGAAATCT AGAAGAGAAT TTTCTGATTG TCTAAAAACA CATTTGGGAT CTTTGTGCAA TTTTATGGGT

5
10
15
20
25
30
35
40
45

```

          920          940          960          980          1000
TTTTTAAAAA CATGTTTGA ACITGATGAG TAAGATTAAT GGAAGAAATG ATGTTTTTGT GTGCTGTGAA GCATGCTTCG GACACTGGAG AGGTACCAAA
AAAAATTTTT GTACAAAAC TGAAGTACTC ATTCTAATTA CCTTCTTTAC TACAAAAACA CACCACACTT CGTACGAAGC CTGTGACCTC TCCATGGTTT

          1020          1040          1060          1080          1100
AGTGTAAC TA TGAAGCACCA GAACCCAATG TGCCTTCAAG AGAGGCCCTTA GCAGTTGTAC CCAATTCTCT TCTCTTTCTT CTAATTACCT TAATTAATTA
TCACATTGAT ACCTCGTGGT CTGGGGTAC ACGGAAGTTC TCTCCGGAAT CGTCAACATG GGTTAAGAGA AGAGAAAGAA GATTAATGGA ATTAATTAAT

          1120          1140          1160          1180          1200
CTCTCAATTT TTACTTTGAT TTTTAGAGTC AAATGATTA TGTTATAATT TGTATATATC TTCAGGAAC TTAGTAGCCAG CAGGAGTATC TCAAGCTTAA
GAGAGTTAAA AATGAAACTA AAAATCTCAG TTTACTAATT ACAATATTAA ACAGTATATG AAGTCTCTGA ATCATCGGTC GTCCCTCATAG AGTTCGAATT

          1220          1240          1260          1280          1300
GGAGCGTTAT GACGCCITAC AGAGAACCCA AAGGTAAACT AATTAGCTTC TTCAGCTACC TTCAGAGAGT GTTTGTTTTT TTAGTAGATT TTTTGTATGG
CCTCGCAATA CTGCGGAATG TCTCTTGGGT TTCCATTGTA TTAATCGAAG AAGTCGATGG AAGTCTCTCA CAAACAAAAA AATCATCTAA AAAAAGTACC

          1320          1340          1360          1380          1400
TTTTGATGTT GAAATAGGAA TCTGTTGGGA GAAGTACTTG GACCTCTAAG TACAAAGGAG CTTGAGTCAC TTGAGAGACA GCTTGATTCT TCCTTGAAGC
AAAACTACAA CTTTATCCTT AGACAAACCT CTCTTAGAAC CTGGAGATTG ATGTTTCTCT GAACTCAGTG AACTCTCTGT CGAACTAAGA AGGAAGTCTG

          1420          1440          1460          1480          1500
AGATCAGAGC TCTCAGGTA CTACTTTGTT CATCAATATC TTTATACACT GATCTATTTC CATAGTAAGA TTAATTTTGG TGTTTAATTC TGCAGACACA
TCTAGTCTCG AGAGTCCCAT GATGAAACAA GTAGTTATAG AAATATGTGA CTAGATAAAG GTATCATTTCT AATTTAAACC ACAAATTAAG ACGTCTGTGT

          1520          1540          1560          1580          1600
GTTTATGCTT GACCAAGTCA ACGATCTTCA GAGTAAGGTA AATAAGAAA CACTCATTTCT CCTCTCTAAA TTCTCTCATC AAAAGTAATG TAACCAAGAA
CAAAATACGA CTGTCGAGT TGCTAGAAGT CTCAATCCAT TTATTTCTTT GTGAGTAAGA GGAGAGATTT AAGGAGTAGA TTTTCATTAC ATTGGTTCTT

          1620          1640          1660          1680          1700
AACACAAATA TTTGAGCAG GAACGCATGC TGACTGAGAC AAATAAAACT CTAAGACTAA GGGTAATTA TATACATTCT CATATCACCA AATTAATGCA
TTGTGTTTAT AAACCTCGTC CTGCGTACG ACTGACTCTG TTTATTTTGA GATTCTGATT CCCATTAATT ATATGTAAGA GTATAGTGGT TTAATTACGT

          1720          1740          1760          1780          1800
TCACTAAATT TGGTTATAAT GTGTGTGTGT ATATACATAT GTGACAGTTA GCTGATGGGT ATCAGATGCC ACTCCAGCTG AACCTTAACC AAGAAGAGGT
AGTGATTTAA ACCAATATTA CACACACACA TATATGTATA CACTGTCAAT CGACTACCCA TAGTCTACGG TGAGGTCGAC TTGGGATTGG TTCTTCTCCA

          1820          1840          1860          1880          1900
TGATCACTAC GGTGTCATC ATCATCAACA ACAACAACAC TCCCAAGCTT TCTTCCAGCC TTGGAATGT GAAACCATTC TTCAGATCGG GTAACCTTTAG
ACTAGTGATG CCAGCAGTAG TAGTAGTTGT TGTGTTGTGT AGGGTTCGAA AGAAGTCCG AAACCTTACA CTTGGGTAAG AAGTCTAGCC CATTGAAATC

          1920          1940          1960          1980          2000
ACTAGTATAA CCAATTGAT TTGAGTTCTA TTATAAGCTT TTCTTAAGAA AGTATCTCAA ACTACTAAAT TTTATGGAGC AGGTATCAGG GGCAACAAGA
TGATCATATT GGTTAACATA AACTCAAGAT AATATTGAA AAGAATTCCT TCATAGAGTT TGATGATTTA AAATACCTCG TCCATAGTCC CGTGTGTTCT

          2020          2040          2060
TGGAATGGGA GCAGGACCAA GTGTGAATAA TTACATGTTG GGTGGTITAC CTTATGACAC CAACTCTATT
ACCTTACCCT CGTCTGTTT CACACTTATT AATGTACAAC CCAACCAATG GAATACTGTG GTTGAGATAA

```

SEQ ID NO:44

Arabidopsis AGL20 genomic sequence

55
60
65
70
75

```

          -2981          -2961          -2941          -2921          -2901
GAAAAAAGAA ACACCTAAG AAGTGAATAT AATAGGCATA TACATATGAG GAAAAATGAA ACAAAAGGAG CGAAAAATAG ATTTAACCTA AAAGAGGAAG
CTTTTTTTTT TGTGGATTTC TTCACTTATA TTATCCGTAT ATGTATATCT CTTTACTATT TGTTTTCTCT GCTTTTATAT TAAATTGGAT TTCTCTCTTC

          -2881          -2861          -2841          -2821          -2801
TAAAGAGGTT ATAAGAGGTA AGAAAAGTAG GACCATATATA TAGCTATATT GTAGAATTTT ATTAATTTGGA GATATGGCAA TTTTGTGTAG GGTCCCATGA
ATTTCCTCAA TATTCTCCAT TCTTTTCATC CTGTATATAT ATCGATATAA CATCTTAAAA TAATAAACCT CTATACCGTT AAAAACACTC CCAGGTACTT

          -2781          -2761          -2741          -2721          -2701
AGACTAAAGT GTGAGACACG ATTTATCTTT GTAATTAATA AAATAATAAA TATATTATTA TTGCTCTGGG ATTTTTCGAT TGATGAGAAA AAGTAAGAGG
TCTGATTCCA CACCTCTGTC TAAATAGAAA CATTAATTAT TTTATTTATT ATATAATAAT AACAGAGCCC TAAAAAGCTA ACTACTCTTT TTCACTTCTC

          -2681          -2661          -2641          -2621          -2601
TGCGTTTTCG AATTATCAAT GGCTAACGTT TGTCAGTGAC TGTACGGAGC ACGTTGATGT ATTTCTAATA TTGACTCTTT TTTTCCACC CTTATTTCTC
ACGCAAAAGC TTAATAGTAA CCGATTGCAA ACATGCACCTG ACATGCCTGC TGCAACTACA TAAGATTTAT AACATGAGAA AAAAGGGTGG GAATAAGAG

          -2581          -2561          -2541          -2521          -2501
TAATTCCTGT ACATTAAACC CAAACTAATT TTACAAACAC ATTGGTGTTT AATCATTGTG AAATTTTGAT TTATCTAAAA TACACTTAT ATGTTATGAT
ATTAAAGACA TGTAATTGGG GTTTGATTAA AATGTTTGTG TAACCACAAA TTAGTAACAC TTTAAAACTA AATGATTTT ATGTGAAATA TACAATATA

          -2481          -2461          -2441          -2421          -2401
TTTGCATGAG CTTATGACTG GTAACTCAT GAGATTTCCA TATCACCATG TTGGAAGTTA CTAACCATAC ATCTTTTAAA TGCAAAATCA CATCATTCCT
AAACGTACTC GAATCTGAC CATTTGAGTA CTCTAAAGGT ATAGTGTATC AACCTTCAAT GATTGGTATG TAGAAATTTT ACGTTTAAGT GTAGTAAGGA

```

		-2381		-2361		-2341		-2321		-2301	
		AGACTGCTAG	ACAGACATGT	ACTTACTTAT	ACAAGGTTTT	TCTAATCTAA	TGGCAACAAA	GAAACTTGTG	ACTAAACGCA	TACGTATCTC	TCATATAGTG
		TCTGACGATC	TGTCGTGACA	TGAATGAATA	TGTTCCAAAA	AGATTAGATT	ACCGTTGTTT	CTTTGAACAC	TGATTTGCGT	ATGCATAGAG	AGTATATCAC
5			-2281		-2261		-2241		-2221		-2201
		TAGACTAGAA	CTCTACGTAT	CTCTCATATA	GTCAATTTTT	TAAAAAAATT	ATACTTTGGG	ATCTCGAAGC	GAAAAATTGA	TTAGTTTATA	TGATTATGTA
		ATCTGATCTT	GAGATGCATA	GAGAGTATAT	CAGTATAAAA	ATTTTTTTAA	TATGAAACCC	TAGAGCTTCG	CTTTTAATCT	AATCAAAATAT	ACTAATACAT
10			-2181		-2161		-2141		-2121		-2101
		CAAAAAAAT	CGGATATTAC	TACCACCTAA	AAATAAATTGT	AGTGGTCAAT	CATATCTAAA	ATTAATCGCA	GTGAACAAAA	ACCTGAAGCA	TAGCCTGGTT
		GTTTTTTTTA	GCCTATAATG	ATGGTGAATT	TTTATTAACT	TCACCAGTTA	GTATAGATTT	TAATTAGCGT	CACCTGTTTT	TGGACTTCGT	ATCGGACCAC
15			-2081		-2061		-2041		-2021		-2001
		CTATCTTACT	TTCGATGTGA	CACATTACTA	ACACGATTGT	TTTAATCTAT	AGGACGAATC	CTTTAAGTAA	TGTATAGTTG	GTTCAGTTAC	GTGATGATCA
		GATAGAATGA	AAGCTACACT	GTGTAATGAT	TGTGCTAACA	AAATTAGATA	TCCTGCTTAG	GAAATTCATT	ACATATCAAC	CAAGTCAATG	CAATCTATGA
20			-1981		-1961		-1941		-1921		-1901
		TTTTGTTTTG	GATTGTGCTC	AACCAGTTAA	GAAGTGATCG	TATTTACTAG	TGGTATACGA	TGATGTTTTCT	TTAAATCTGA	ATTGGGTCTA	CAAAATACAT
		AAAACAAAAC	CTAAACACAG	TTGGTCAATT	CTTCACATAG	ATAAATGATC	ACCATATGCT	ACTACAAAGA	AAATTAGACT	TAAACCCAGAT	GTTTTATGTA
25			-1881		-1861		-1841		-1821		-1801
		AACTAAACTT	CAAAACCGGT	TTATACTTTA	TACAAACACG	AAAAATAAAA	GATAGAGACA	ATTCCACCAG	GAAGATGTGT	ATTTATATAA	AAATTATCCA
		TTGATTAGAA	GTGTGGCCCA	AATATGAAAT	ATGTTGTGTC	TTTTATATTT	CTATCTCTGT	TAAAGTGTCT	CTTCTACACA	TAAATATATT	TTTAATAGGT
30			-1781		-1761		-1741		-1721		-1701
		TACAGATTTT	CGGACCTATC	TGTTTGATAT	TTAATATATA	TAAATACGTT	AACATAITTC	ACCAGAGAAG	ATGTGTAITT	TTGCAAAATA	TTAGTTTGTG
		ATGCTCAAAA	GCCTGGATAG	ACAAACTATA	AATTATATAT	ATTTATGCAA	TTGTATAAAG	TGGTCTCTTC	TACACATAAA	AAAGCTTTAT	AATCAAAAC
35			-1681		-1661		-1641		-1621		-1601
		TGGTCTCTCT	CCCGATATAT	ATAAAAGATC	ATTAGATATC	GATTAAACAAT	TTTATCTCCA	AAAAAGGATA	TTTTTTTGGT	GCCACTAGCT	AGACAAGACG
		ACCAGGAGGA	GGGCTATATC	TATTTTCTAG	TAACTCTATG	CTAAATTGTTA	AAATAGAGGT	TTTTTCTCAT	AAAAAAACCA	CGGTGATCGA	TCTGTTCTGC
40			-1581		-1561		-1541		-1521		-1501
		TTGATTAAGC	TGAATTTATTA	TTGGATTCTC	AAGTTACGTT	TTCTTTAGTA	ATCCGAGGGA	CCAAAAATAG	CAAAATGCTT	TTTACAGACG	TCGCTACTTA
		AAGCTATTCT	ACTTAATTAAT	AACCTAAAGA	TTCAATGCAA	AAGAAATCAT	TAGGCTCCCT	GGTTTTTATC	GTTTTACGGAG	AAATCTGTGC	AGCGATGAAT
45			-1481		-1461		-1441		-1421		-1401
		ACGCCATTGC	CCCATGTGCT	CTGTACTAGC	CTCCAAATAT	TTGGATTAAAT	GGTCACCTAG	GTAATGAGGA	AATTTAGTAG	TTTTGTAATG	TGGTTTTGTC
		TGCGGTAACG	GGGTAAACAG	GACATGATCG	GAGGTTTTATA	AACTTAATTA	CCAGTGAATC	CATTACTCCT	TTAACATCAT	AAAAACATTAC	ACCAAAACAG
50			-1381		-1361		-1341		-1321		-1301
		CAACTTATAA	AAACTTTACAA	TTGCAAGTAA	TTAATTATTC	ACATGGAGAT	GTAAGATTAT	GTCAATAAAC	TAAAAACACA	ATTTAAGAAC	AACAATAAGA
		GTTGAATATT	TTTGAATGTT	AACGTTCAAT	AATTAATAAG	TGTACCTCTA	CATTCTAATA	CAGTATATTG	ATTTTGTGTT	TAAATCTCTG	TTGTTATTCT
55			-1281		-1261		-1241		-1221		-1201
		AACAATGGAC	AAACAAGCAT	AGAAAAATATA	CAAAATCAAGT	GAATTTTATC	TGTTGGGATG	GAAAGATATT	ATAAAAACCA	ATTAAGTTGT	ATATAGTTGT
		TTGTTACCTG	TTTGTTCGTA	TCCTTTATAT	GTTTAGTTTA	CTTAAATATG	ACAACCTTAC	CTTTCTATAA	TATTTTAAAC	TAATTTTGGT	TATATCAACA
60			-1181		-1161		-1141		-1121		-1101
		ATTACTCACA	GGTAAGAAAA	AACGATATTC	TTATTTTCCA	TATCAATTAC	AAGTGGGGGC	ATATAGGTAC	GAGAGAGTGT	TTGTGTCCAC	ATTAAAAACA
		TAATGAGTGT	CCATCTCTTT	TTGCTATAAG	AATAAAAAAT	ATAGTTAATG	TTCAACCCCG	TATATCCATG	CTCTCTCACA	AACACAGGTG	TAATTTTTGT
65			-1081		-1061		-1041		-1021		-1001
		AAAAAAGATT	TTTGTTAGAA	GAAATTTAAT	AAAAATAATT	TGACAGGCAT	TTCCATCCAA	CTAGATATTT	ATGGGAGGGA	AAAAAGATGT	TATGTAAAAA
		TTTTTTCTAA	AAACAATCTT	CTTTAAATTA	TTTTTATATA	ACTGTCCGTA	AAGGTAGGTT	GATCTATAAA	TACCTCCCTC	TTTTCTACAC	ATACATTTTT
70			-981		-961		-941		-921		-901
		TGTCCATATG	TATCAAAAAA	TGCTATTTTT	GGTCTTTCTT	AAGGCTTTTT	TCCAAAAATA	GTAAGAGGAT	AGGTTTCAAG	CGTCCATCAT	ATTTGCGACA
		ACAGGTATAC	ATAGTTTTAT	ACGATAAAAA	CCAGAAAGAA	TTCCGAAAAA	AGGTTTTATT	CATTTCCTAC	TCCAAAGTTC	GCAGGTAGTA	TAAACGCTGT
75			-881		-861		-841		-821		-801
		CATATGACTG	ACTAATTTAGC	TCCTCCCTCT	TTCTTTCTCT	TATTTTATTA	TCTTTCTCCA	AGAAATAAAA	TAGAAAAAGAA	AATATATATG	GTTCACAAAA
		GTATACTGAC	TGATAAATCG	AGGAGGGAGA	AAGAAAGAGA	ATAAAATAAT	AGAAAGAGGT	TCITTTATTT	ATCTTTTCTT	TTATATATAC	CAAAAGTTTT
80			-781		-761		-741		-721		-701
		CACCAATTACC	ATAACTACAA	CGAGAAGAGG	ATCTTTTTTT	AGGAGAAAAA	CAGAGAGAGA	AGAGACGAGT	GTGTGAAGTT	TTTTTGTCTT	TTGTTTCTTT
		GTGGTAATGG	TATTTAGTGT	GCTCTTCTCC	TAGAAAAAAT	TCCTCTTTTC	GTCTCTCTCT	TCTCTGCTCA	CACACTTCAA	AAAAACAGAA	AACCAAGAAA
85			-681		-661		-641		-621		-601
		TATTACACAC	AAATAGATGA	AACGAGGAAA	GCTACTTCTT	TTGCTACTTC	CATAAAAGAG	TTCTTCTCTT	CGCAGAGAAT	CAACTTTGAT	CATCTTCTTC
		ATAATGTGTG	TTTATCTACT	TTGCTCCTTT	CGATGAAGAA	AACGATGAAG	GTATTTTTCC	AAGAAGGAAA	GCCTCTCTTA	GTGGAACATA	GTAGAAGAAG
90			-581		-561		-541		-521		-501
		CTTCTCTTTC	TTTCTTCTTC	TCCCTCCAGT	AATGCTTATA	TAGTCTCCTC	CTATATCTCT	ACCTATACAT	ACACAAACCC	TTTATCTCTG	AAAGCTTCTC
		GAAGAGAAAG	AAAGAAGAAG	AGGGAGGTCA	TTACGAATAT	ATCAGAGGAG	GATATAGAGA	TGGATATGTA	TGTGTTTTGG	AAATAGGAGC	TTTCGAAGGA
95			-481		-461		-441		-421		-401
		CCTGGTTAGG	TTTTTATCAA	ACCCTTTTAT	CCAATCGGTA	AGATCTCTTC	GTCAATGATC	TTTCTTTTTT	CTTTTGTCTT	GTACTCTGAT	GGATCTATAA
		GGACCAATCC	AAAAATAGTT	TGGGAAAAAT	GGTTAGCCAT	TCTAGAGAAG	CAGTACTAGA	AAAGAAAAAA	GAAACAGAAA	CATGAGACTA	CCTAGATATT
100			-381		-361		-341		-321		-301
		ACTTATATGG	GTTTGGTTTC	ATTTGGTTTC	ATTTGATGTC	TTTGGTTTCT	TTGTCTCTAA	TCTCATGAAA	GGAGGTTTGA	TCCTTCAATT	AAACCGATAA
		TGAATATACC	CAAAACCAAG	TAAACCAAGC	TAAACTACAC	AAACCAAGAA	AACAGGATTT	AGAGTACTTT	CCTCCAACGT	AGGAAGTTAA	TTTGGCTATT
105			-281		-261		-241		-221		-201
		CAAAAGTTTC	CATTACAGAC	TTATAGATCA	GATACCTTTG	ATTGTTTTGC	TTTTTGGGTA	CTTAATCTTT	CGTTGACTTC	ATCAGTCTTC	TCCACCCCAA
		GTTTTCAAGG	GTAAATGCTG	AATATCTAGT	CTATGAAATC	TAACAAAACG	AAAAACCCAT	GAATTAGAAA	GCAACTGAAG	TAGTCAGAAG	AGGGTGGGTT
110			-181		-161		-141		-121		-101
		ACAAAAAAGT	CATATTTTGA	TCATATCTTC	ATTTTTTTAA	CCTACTCTCT	TTGATTTCTA	TATGAAATGG	GTGTTTTTAT	GTGTGTGACT	AATCTTGTTA
		TGTTTTTTCA	GTATAAAGCT	AGTATAGAAG	TAAAAAAATT	GGATGAGAGA	AACTAAGTAT	ATACTTTACC	CAACAAAATA	CACACACTGA	TTAGAACAAT
115			-81		-61		-41		-21		-1
		TTGAGGTGGT	TGCACCATTT	ATCTACCGTT	TTCTTCAATT	TTTGAAAAAA	TAATTTTATT	TTTTTCTGTG	GTGCAAGGGA	AATTAACATA	AGAAGAAGAT
		AACTCCACCA	ACGTGGTAAC	TAGATGGCAA	AAGAAGTTAA	AAACTTTTTT	ATTAATAATA	AAAAAGACA	CACGTTCCCT	TTAATTGATT	TCTTCTTCTA
120			20		40		60		80		100
		ATGGTGAGGG	GCAAAACTCA	GATGAAGAGA	ATAGAGAATG	CAACAAGCAG	ACAAGTGACT	TTCTCCAAAA	GAAGGAATGG	TTTGTTGAAG	AAAGCCTTTG
		TACCACCTCCC	CGTTTTGAGT	CTACTTCTCT	TATCTCTTAC	GTGTTTCTGC	TGTTCACTGA	AAGAGGTTTT	CTTCTTACC	AAACAACCTC	TTTCGGAAC
125			120		140		160		180		200
		AGCTCTCAGT	GCTTTGTGAT	GCTGAAGTTT	CTCTTATCAT	CTTCTCTCCT	AAAGGCAAAAC	TTTATGAATT	CGCCAGCTTC	AAGTACGTTT	TTTTTGTCTT

TCGAGAGTCA CGAAACACTA CGACTTCAAA GAGAATAGTA GAAGAGAGGA TTTCCGTTTG AAATACTTAA GCGGTGAGG TTCATGCAAG AAAAACAGAA

220 240 260 280 300

TCTTACAAAT CATCCATAGA AAGAGAGAGA GAGAGAGATC TCATTAACCT CTCTATTTGT ATCTTAATTT TTTTGGTTT ATATATGGAT TTGATTGGCC

AGAATGTTTA GTAGGTATCT TTCTCTCTCT CTCTCTCTAG AGTAATTGGA GAGATAAACA TAGAATTAAT AAAAACCAAA TATATACCTA AACTAACCCG

320 340 360 380 400

TTTTGTGGAA TCACATCTCT TTGACGTTTG CTTTGAGAGG TGTGTTTAAA TGAGTTTCTT GGTTCCTGCA AAATTAGGGC TATTATTAAA GTAGTATCAA

AAAACACCTT AGTGTAGAGA AACTGCAAA GAAACTCTCC ACACAAATTT ACTCAAGAA CCAAAGACGT TTTAATCCCG ATAATAATTT CATCATAGTT

420 440 460 480 500

GTACATATAC CCTCTTATTT ATTGTTTTTT TATTTCCGCT AGTATATCAT CTGTGTTAAT CATCTGTCTC TCTCTTCTC AATTAGTTTC TCAAGTTATG

CATGTATATG GGAGAAATAA TAACAAAAAA ATAAAGGCGA TCATATAGTA GAACAAATTA GTAGACAGAG AGAGAAAGAG TTAATCAAG AGTTCAATAC

520 540 560 580 600

ATATAAATAA AATGTGCTCT TTCGTAGCCA ATTTACACTT GTTATATATT TGATCTTCTT AGAGATCATG ATCACATAGT ATTAATAAAA CAACTTTCAA

TATATTTATT TTACACGAGA AAGCATCGGT TAAATGTGAA CAATATATAA ACTAGAAGAA TCTCTAGTAC TAGTGTATCA TAATTATTTT GTTGAAGTT

620 640 660 680 700

TTAGTATTCT TTTGGTTTGA ACTAATCTTT GTCTGTGTTT TGCTTTAAGC AAAACATGTT CTAAGTGTAT ATTAGGAAGT TGTTCATCA

AATCATAAGA AAACCAAACT TGATTAGAAA CAGAACAATA ACGAATTTG TTTTGTACAA CAAGATTAAA GATTCACTAC TAATCCTTCA ACAAGTAGT

720 740 760 780 800

TTCTGTAITT ATTAATCCCT CATGCTTCAT TTCATGCTCA TTCTTAATTT AGTTCAATTT GTTTGAATAT TTGTTCTGTA TTTTGACATA GAAACTCAAA

AAGGACTAAA TAATTAGGGA GTACGAAGTA AAGTACGAGT AAGGATTAAA TCAAGTTAAA CAAACTTATA AAACAGGACT AAAACTGTAT CTTTGAGTTT

820 840 860 880 900

GCTAGCTAGC CAAACCTAAA TGTGATTGTT TTTTGAGAAT CAAAAGAGTT TTATCTTGTA CTGTTAGGTA GTAGGGAAGC CAAACTTACT TTTGATTCAA

CGATCGATCG GTTTGGAATT ACAACTAACA AAAACTCTTA GTTTTCTCAA AATAGAACAT GACAACTCAT CATCCCTTTG GTTGAATGA AAACACTTAA

920 940 960 980 1000

CATTACTTCT GTAAATGAAA ATGCCAGCTT TTGATCAGAT GTTTCAGACA TTTGGTCCAT TTGGGAAAGT ACTTCTTCT CTCGAACCTA CTAATATATA

GTAATGAAGA CATTTACTTT TACCGTCGAA AACTAGTCTA CAAAGTCTGT AAACCAAGTA AACCTTTCA TGAAGAAAGA GAGCTTGGAT GATTTATATT

1020 1040 1060 1080 1100

AGATAAGACC TCACATGTTT TTGATTTTTCT AAAATAGGGG GAAAAAGTAC AAGACTTTTC AAGCTATGTC CTGATTGAAG TCTAGTATA TCTTCAATAA

TCTATCTCGG AGTGTACAAA AACTAAAAAG TTTTATCCCC CTTTTCATG TTTCTGAAAAG TTCGATACAG GAACATAATC AGATCACTAT AGAAGTTATT

1120 1140 1160 1180 1200

GAAATGTTT GAGAACACCA TTGGGATCTA AAATTGATCT CTGATGATTT ACTTTAATGT TCCTTAATATA TAGTTTATG ACAGTATGCA AGATACCATA

CTTTACAAAA CTCTTGTGTT AACCTAGAT TTAACCTAGA GACTACTAAA TGAAATTACA AGGTTAATAT ATACAAAAAC TGTCATACGT TCTATGGTAT

1220 1240 1260 1280 1300

GATCGTTATC TGAGGACATC TAAGGATCGA GTCAGCACCA AACCAGTTTC TGAAGAAAAT ATGCAGGTTT ATTCTTTG CTAATATATA CTATATAGT

CTAGCAATAG ACTCCGTATG ATTCCTAGCT CAGTCGTGTT TTGGCCAAAG ACTTCTTTTA TACGTCACAA TAAGAAATAC TAGAAGAACG GATATATAGT

1320 1340 1360 1380 1400

ATTCTTGCTA ATTAATGTTT TTAATATATA ATATCAAGA GCGGTAATGA ATATAACCAC AATATGTATA TAATCTCAAG GTCCACAGAT CAAGTCACAT

TAAGAACGAT TAATTATGAA AATGATATAT TATAGTTTCT CGCCATTACT TATATTGGTG TTATACATAT ATTAGAGTTC CAGTGTCTTA GTTCAGTGTA

1420 1440 1460 1480 1500

ATTTATAATT AGGATATATA TGATCATGCA ATAACTTTTC TGTGATATA CCAACAGCAT TTGAAATATG AAGCAGCAAA CATGATGAAG AAAATTTGAA

TAAATATTA TCCATATATA ACATGTACGT TATTGTAAG ACACATATAT GGTGTGCTGA AACTTTATAC TTCGTCTGTT GTACTACTTC TTTTAACTTG

1520 1540 1560 1580 1600

AATCGAAGC TTCTAAACGG TTTGTGATAT ATACATATAT ACAAACACAT TATTCATCAC TTGTATATAT CTAATTCATG ATGCATAGGA GAGTTTATC

TTGAGCTTCG AAGATTGTC AACAACATATA TATGTATATA TGTTTGTGTA ATAAGTAGTG AACATATATA GATAAAGTAC TACGTATCTC CTCAAACTAG

1620 1640 1660 1680 1700

AATTAGTGTT TTGTTTGTGT AATCAGTAAA CTCTTGGGAG AAGGCATAGG AACATGCTCA ATCGAGGAGC TGCAACAGAT TGAGCAACAG CTTGAGAAAA

TTAATCACAA AACAAAAACA TTAGTCATTT GAGAACCCCT TTCCGTATCC TTGTACGAGT TAGCTCCTCG ACGTTGTCTA ACTCGTTGTC GAACTCTTTT

1720 1740 1760 1780 1800

GTGTCAAATG TATTCGAGCA AGAAAGGTAT GTGTATATAT TTATCTGTTA TATCTCCACA TTATAAGTAT TGTTCGATC ATCTCTGAA ACCACTATA

CACAGTTTAC ATAAGCTCGT TCTTTCCATA CACATATATA AATAGACAAT ATAGAGGTGT AATATTCATA ACAAGCTTAG TAGAAGACTT TGGTAGTAT

1820 1840 1860 1880 1900

ATTATAACTC AATTCTCAT CTCTTTTAGA CTCAAGTGT TAAGGAACAA ATTGAGCAGC TCAAGCAAAA GGTAAAGTAG TTTTATAGG TGTATATAAA

TAATATTGAG TTAAAGAGTA GAGAAAACT GAGTTCACAA ATTCCTTGT TAACCTGTCG AGTTCGTTTT CCATTTCATC AAAAATACTC ACATATATTT

1920 1940 1960 1980 2000

CAGATATAAG TATGTATGCA AATTGTGTAA TATTCAGAT AAGTAAGCCT CTGTGCTCTG CTTTITACAA ATTGGAATCT AAAACTTTTG CAGGAGAAAG

GTCTATATTC ATACATACGT TTAACACATT ATAAGGTTCA TTCAITCGGA GAACACGAA CCAATAGAGT TTAACCTTAG TTTTGAAAAA GTCCTCTTTC

2020 2040 2060 2080 2100

CTCTAGCTGC AGAAACGAG AAGCTCTCTG AAAAGGTATA ATATATTCTT ATGGGTCTCA AGTTAGGGTT GCACATTCTG TTTTATTATC GGTAAAGATA

GAGATCGAGC TCTTTGCTC TTGAGAGAC TTTTCCATAT TATATAAGAA TACCCAGAGT TCAATCCCAA CGTGTAAAGCA AAAAATAAG CCAATTCTAT

2120 2140 2160 2180 2200

AGAAAGTTGG GGTCTCTTTT GGGGGTTATT AGGTTAGGAG AGTCCTTACT AGTTTCTCTT GGTATCTTTC AATCATCAAC CTTCTTTAAT TTATGTATTG

TCTTTCAACC CCAAGAAAAA CCCCCAATAA TCCAATCTCT TCAGGAATGA TCAAAAAGAA CCAATAGAGT TTAGTAGTTG GAAGAAATTA AATACATAAC

2220 2240 2260 2280 2300

TTCTATATAT CTCTCAATTT GCATCTATTA ATTTTGTGTA ATAATTCTAT TTGAATGCAG AACTTACGTC TGGGATCTC ATGAAAGCGA AGTTTGGTCA AATAAGAAATC

AAGATATATA GAAGATTAAA CGTAGATAAT TAAACACAT TATTAAGATA AACTTACGTC ACCCTAGAG TACTTTCGCT TCAATCCAGT TTATCTTAG

2320 2340 2360 2380

AAGAAAGTAC TGAAGAGGT GATGAAGAGA GTAGCCCAAG TTCTGAAGTA GAGACGCAAT TGTTCATTGG GTTACCTTGT TCTTCAAGAA AG

TTCTTTCATG ACCTTCTCCA CTACTTCTCT CATCGGGTTC AAGACTTCAT CTCTGCGTTA ACAAAGTAAC CAATGGAACA AGAAGTTCTT TC

SEQ ID NO:45

Arabidopsis AGL22 genomic sequence

2981 2961 2941 2921 2901

TACAAGTCAT CGCGCCGCTG GTCATTTTCA GGATCCGGCG AGAAACTGAA CCAAAAATAT ACTTATTTTA CTGTAAGGA -2901

ATGTTTCAGTA GCGGCGCGAG CAGTAAAGT CCTAGCGCGC TCCTTGACTT GGTTTTATTA TGAATAAAT GAGCATTCCT TTTAAACCCG GATTATTTTC

2881 2861 2841 2821 2801

CCCAATAATA ATAAAAAGCC CATTAGGGAC TCCGCTTTAT GATAACGGTG ACTGTAGTTT CCTGTAGTGT TCAGAGAGAG TGTGTAGTGT AGGAGCTGTG

GGGTATTAT TATTTTTCGG GTAATCCCTG AGGCGAAATA CTATTGCCAC TGACATCAAA GGAACACAC ACATCTCTCT ACACATCACA TCCCTGACAC

		-2781		-2761		-2741		-2721		-2701	
		TAGAAAGAAA	GAAGCCTAAA	ATGGCTAAAA	GGTTAGGTGC	AATGTTTCAT	TAGAGAGGCT	TGGAAGCTGT	AAGGGAAGG	TCACGAGTCG	TCTACTCATA
5		ATCTTTCTTT	CTTCGGATTT	TACCGATTTT	CCAAATCCAGC	TTACAAAGTA	ATCTCTCCGA	ACCTTGACAA	TTCCCTTTCC	AGTGCTCAGC	AGATGAGTAT
		-2681		-2661		-2641		-2621			
		AAAACCTCTGA	CACCTTTGACC	AATCAAAACT	CAAGAGCCTC	ACCAGTTGTG	TCACGTGCGC	CTCTAAACAC	TATTTCAATT	CAAAATATAA	TGATTTCATC
10		TTTGTGAGACT	GTGAAACTGG	TTAGTTTGTG	GTTCCTGGAG	TGGTCAACAC	AGTGACACGG	GAGATTTGTG	ATAAGTTAAA	GTTTATATTT	ACTAAGTACG
		-2581		-2561		-2541		-2521		-2501	
		GGTTCCAAAC	GCCAAATGAT	GGATGTTCTA	CCAAATTTAA	TCTACTTTTA	CCAAACCATG	ACAAATATGA	ATAACCATTA	CTTGATAATA	ATTTTGTGAG
		CCAAGGTTTG	CGGTTAACTA	CCTACAAGAT	GGTTTAAATT	AGATGAAAA	GGTTTGTGAC	TGTTTATACT	TATTTTGAAT	GAACATTTAT	TAAACACTC
15		-2481		-2461		-2441		-2421		-2401	
		TGAACAAACT	TTTTTTTTTT	CGAAACCAAA	CCAAGCTGAA	AAAAACTCAA	CGATTTTCTT	TGTTTAAAA	ACGTTAGAAA	GGAATATGTA	TTATGCCGAA
		ACTTGTTTGA	AAAAAAA	GCTTTGGTTT	GGTTCGACTT	TTTTTTCAGT	GCTTAAAGAA	ACAAATTTTA	TGCAATCTTT	CCTTATACAT	AATACGGCTT
		-2381		-2361		-2341		-2321		-2301	
		ATAAGTAATA	TCGATCAGGC	CACCTCTCTT	ATAGTTATTC	TCCTAGCAAC	TTTAACCACT	AGAAGGTTTT	GTTCCTTAGT	GTTCCTTAAT	ATACGTCTAT
20		TATTCATTAT	AGCTAGTCCG	GTGGAGAGAA	TATCAATAAG	AGGATCGTTG	AAATTTGGTA	TCTTCCAAAA	CAAAAGATCA	CAAAAGATTA	TATCGAGTAG
		-2281		-2261		-2241		-2221		-2201	
		AAAATTTTCA	AAAAATACTA	CATTTTGTGT	TTAAAAACTT	CCATAATTCC	ATTACTCGTA	GAACACAAAC	GCAAAACATA	TTAATATTTT	GTGTCTCAAC
25		TTTTAAAGTT	TTTTTATGAT	GTAAAAACAA	AATTTTGGAA	GGTATTAAGG	TAATGAGCAT	CTTGCTTTTG	CGTTTGGTAT	AATTATAAAA	CAACAGTTGT
		-2181		-2161		-2141		-2121		-2101	
		AAAATTTCAA	ATATATAATC	AACATATATT	GCTTGATTAC	CCAATTAGAT	AGAAAGAGGT	TAAAGAGAA	AAGAAAGAG	TTTACAGTAA	ATTAACGCCAA
		TTTTAAAGTT	TAATATAAAG	TTGATATAAA	CGAACTAATG	GGTATTAAGG	TCTTTTCTCA	ATTTCTTCTT	TTCTTTTCTC	AAATGTCAAT	TAATTCGGTT
30		-2081		-2061		-2041		-2021		-2001	
		ACCATAATTA	TATTTAACAC	CGTATTAATC	ACATCAACCA	TATGACTTTT	TTACCGTTTG	CAACTTCATA	ATTCATATAG	TATCATATAA	AATTCGCAAT
		TGGTATTAAT	ATAAATTTGT	GCATAATTAG	TGTAGTTGGT	ATACTGAAAA	AATGGCAAAC	GTGGAAGTAT	TAAGTATATC	ATAGTATTAT	TTAAGCGTTA
35		-1981		-1961		-1941		-1921		-1901	
		AATACAACAC	AAGAGTTTGG	TCGGAAGAGT	AAATAACTAT	CAAAATAGGG	GTGAGTGATA	CGAGCCACAT	GTATCTCTGA	AGGGTAGATT	ATTGCAAACT
		TTATGTTGTG	TTCTCAAAGC	AGCCTTTCTCA	TTTATATAGA	GTTTATCCCC	CACCTCACTAT	GCTCGGTGTA	CATAAGAACT	TCCCATCTAA	TAAAGTTTGA
40		-1881		-1861		-1841		-1821		-1801	
		TGGAGTAATA	AAGAGAAGAA	GAATGGGTTT	GTAGTAGTGT	CGTGAGGATG	CTTTATTTGG	GTAAGAACTT	AATTTAGAAA	TAAAATCTGT	TACGGACAAT
		ACCTCATTTAT	TTCTCTTCTT	CTTACCCAAA	CATCATCAAC	GCACCTCATA	GAAATAAACC	CATTTTGAAA	TTAAATCTTT	ATTTTAAAGC	ATGCGCTGTA
45		-1781		-1761		-1741		-1721		-1701	
		GGATCGGTGC	CCAATCAGAT	TTCTTGTGTG	TGCTTCGGGT	CTGGTTTGTG	GTCCCTTTGA	AAAAATTTAG	TGGTGCAGAC	TTTTTATTTT	ACTCTGGCTC
		CCTAGCACAG	GGTTAGTCTA	AAGAACACCG	ACGAAGCCCA	GACCAAAACC	CAGGGAACCT	TTTTAAATCT	ACCAGCTGTG	AAAAATAAAA	TGAGACCCAG
50		-1681		-1661		-1641		-1621		-1601	
		GTGCCTCGAG	GGTCCCTCTA	TTCACTGTTT	CTTCGTATGA	AGGTATGCTT	AAACATTATT	TTATTTTATA	AAACCCCTTA	ATTTTATTTT	CTTACCTTTA
		CACGGAGCTC	CCAGGAGAGT	AAGTGACAAA	GAAGCATACT	TCCATACGAA	TTTGTAATAA	AATAAAATTT	TTTGGGAAT	TAAATAAAA	GAATGGAAT
55		-1581		-1561		-1541		-1521		-1501	
		ATCACGGTTT	TGTAATTTGC	TTTTTGTGCT	ATGGAATGAT	GATTGTGGCG	ATTGAAATCA	TATGTTTGGT	TCTGTTGTG	ACGTTGGTGA	AGTATATGTT
		TAGTGCCAAA	ACATTTAAGT	AAAAATCAGA	TACCTTAACT	CTAACACCGC	TAACTTTATG	ATACAAACCA	AGACAAACCA	TGCAACCACT	TCATATACAC
60		-1481		-1461		-1441		-1421		-1401	
		ATTTGTAAAT	TTGAGCTTAT	GTATTAATAA	GTTAAATGAT	AAATAACCTC	GTAAGAAAGT	GATTTCAATT	AAATTTTATT	TTGAGTTACA	TATTCATTTG
		TAAACATTAC	AACCTGTAAT	CATAATTTTA	CAATTTACTA	TTTATTTGAG	CATTTCTTCA	CTAAAGTAAA	TTTAAATATA	AACTCAATGT	ATAAGTTAAC
65		-1381		-1361		-1341		-1321		-1301	
		GTTTTATAAA	AAAATACCTC	AGTGATGATT	GATACCCCCA	TTGTGTGTGT	AATTTGTTACT	GGGATGTAAC	AAAATTTATT	TGTGCATGAC	AAACTTTCCA
		CAAAATATTT	TTTTATGAAG	TCACACTACT	CTATGGGGGT	AACACACACA	TTAAACAATG	CCCTAACTTG	TTTTAAATAT	ACACGTACTG	TTTGAAGGTT
70		-1281		-1261		-1241		-1221		-1201	
		AATTAGTGCA	TAGATGTATA	TTGTATAATG	GACTACATGT	ATCTGAGTAG	ATATGGTTCA	TTAGGTTACA	AACCTCTTTT	TTTTAAGACA	CAATTTTCTG
		TTAATCACGT	ACTTAACATT	AACATATTAC	CTGTATGTAC	TAGACTCATC	TATACCAAGT	ATATCAATGT	TTGGAAGAAA	AAATCTCTGT	GTTAAAGAGC
75		-1181		-1161		-1141		-1121		-1101	
		ACAAGTTATA	TGCCACATGA	TTGACTACTA	AATTTTCAAA	AATTATTGCA	CTAATGCTCT	TGAATTTAAC	AAATATTATT	GTCAATTTCC	AGTTGGATTG
		TGTTCAATAT	ACGGTGTACT	AACCTGATGT	TTAAAGCTT	TTAATAACGT	GATTACAGAA	ACTTTAATTT	TTTAAATATA	CAGTAAAGGC	TCAACCTAAG
80		-1081		-1061		-1041		-1021		-1001	
		TTACAAACCA	AGGCCGAAGT	CACAAACCTA	TTTCTTTTCA	TAAACCAAAA	ACATTGTCTC	CAGAAAAATT	CTGAAATGTC	ATCTTCCCAA	ATGTTTCTAC
		AATGTTGGTT	TCCGGCTTGA	GTGTTTGAAT	AAAGAAAGTC	ATTTTGTGTT	TGTAACAGGA	GTCTTTTATA	GACTTTTACG	TAGAAAGGTT	TACAAAAATG
85		-981		-961		-941		-921		-901	
		ATAAATAAAA	ATAATATACA	GTGTATATTA	TTTGTCTTCT	TCTGAATTTT	GTTATGAGGT	ACCATTACCA	TATAGTACGT	AGATTATCAA	AAATGAAAT
		TATTTATTTT	TATTATAGTT	CAACTATAAT	AAAAACAAGAA	CAATCTTCCA	TGGTAATGGT	ATATCATCTG	TCTAAATGTT	TTTACTTTTA	TTTACTTTTA
90		-881		-861		-841		-821		-801	
		ACGTTGTAGC	CCTTGATGTT	CTTCAGGTCT	TCTAGTTAGT	TTTTGCAGTA	AATACCAACC	AATAGTTTAC	AAGGAGTATA	AGTGAACAAA	CTGAGACAAC
		TGCAACATCG	GGAAGTACAA	GAAGTCCAGA	AGATCAATCA	AAAAAGTCTA	TTATGGTTGG	TTAATCAATG	TTCTCTCATAT	TCACTTGTAT	CATCTGTTTG
95		-781		-761		-741		-721		-701	
		TCATTTTATG	CTTCCCTATA	AAAAGAAATT	CCCACTGACG	CCAAACACAC	ACTTCTCTTC	TCTCTCTCAT	CTCATTTGGG	ACCTTATAAT	CCTATTACCT
		AGTAAATATC	GAAGGGATAT	TTTTCTTTAA	GGGGTGAAGT	GGTTTTGTGT	TGAAGAGAAG	AGAGAGAGAG	GAGTAACCTC	AGATTTTATA	GCTATATGGA
100		-681		-661		-641		-621		-601	
		CACCATATCC	AATAACCAAC	ACACACAGAC	CAATATCCAA	AAAAAAACTT	AAAACTAAAA	ATATAATATA	TATCGTTTTC	TTTCCAAAAA	TAATCATTTA
		GTGGTATAGG	TTATTTGGTG	TGTGTGCTGG	GTATATGGTT	TTTTTTTGTG	TTTGTATTTT	TATATTATAT	ATAGCAAAAG	AAAGGTTTTT	ATTAGTAAAT
105		-581		-561		-541		-521		-501	
		AGAAACCCCA	TCATCTTGAT	AGTATTATAA	AATTAATATA	CTCTCTCCCT	AAAAATATCT	ATCCTTCAAC	AATCAAAACC	TTCTCATGTC	TTCTCTCTCT
		TCTTTGGGTT	AGTAGAAGTA	TCATAATATT	TTAATTTATT	GGAGAGGGAC	TTTATATAGC	TAGGAAGTGG	TTAGTTTGGG	AAGAGTACAG	AAGAGAGAGG
110		-481		-461		-441		-421		-401	
		CTCGACCTTT	GAGGTGGAAA	ATTAAATATA	TTCCCTTAGC	TTTTTTTCTC	CTTTAGTTTT	CTTCTCTCTC	TTGAGTTTTT	TTTCTTTTGA	TCCTCTCTAA
		GAGCTGGAAA	CTCCACCTTT	TAATTTATAT	AAGGGAATCG	AAAAAAGAG	GAAATCAAAA	GAAGAAGAAG	AACCTCAAAA	AAAGAAAAC	AGGAGAGATT
115		-381		-361		-341		-321		-301	
		TTTCTTTGTT	GATTCATCGA	CTGATCTCAA	TTCTTCTCAC	AAAAAGCTGA	GTGTGTTCTT	TCCTTCAAA	CTTTCAAAAA	CTAGGGTTTT	TACTGTCTTG
		AAAGGAACAA	CTAAGTAGCT	GATCTAGATT	AAGAAGAGTG	TTTTCTGACT	CACACAAGAA	AGAAAGTTTA	GAAAGTTTTT	GATCCCAAAA	ATGACAGAAC
120		-281		-261		-241		-221		-201	

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

AAATCATATT TATTCTTCTA AATTAGCAAA AAAGAACAGC ATTTACTTTC CATTTTCAGTC GTCTTGTAC TCTCTCTCTC TTCTTTAAAG TCTCCCTTTT
TTTAGTATAA ATAAGAAGAT TTAATTCGTT TTCTTTGTGC TAAATGAAAG GTAAAGTCAG CAGAACAGTG AGAGAGAGAG AAGAAATTTT AGAGGGAAAA

-181 -161 -141 -121 -101
TAGCAAAAT TCTCTCTCTC ACAAATTTA TTCTCTCTGG CTCTCTCTCC ATCTCTCTCT TTTACTCTCT CTTTAATCAT CTCTCATCTT
ATCGTTTTTA AGAGAGAGAG TGTTTTAAAT AAAGGAGACC GAAGAAGAAG GAGGAGGAGG TAGAGAAGAG AAATGAGAGA GAAATTAGTA GAGAGTAAAG

-81 -61 -41 -21 -1
TGAATCTTGA TCCATCAAAA TCAATCCCGT TCTCGAAAGA TCCATTAATA TCAAAACCTA AGCTCTCTCT CTTGCTTTCTA GGGTTTTTTT GTTCGTGTGT
ACTTAGAACT AGGTAGTTTT AGTTAGGGCA AGAGCTTTCT AGGTAATTTT AGTTTTGGAT TCGAGAGAGA GAACGAAGAT CCCAAAAAAA CAAGCAACAC

20 40 60 80 100
ATGCGGAGAG AAAAGATTCA GATCAGGAAG ATCGACAACG CAACGGCGAG ACAAGTGACG TTTTCGAAAC GAAGAAGAGG GCTTTTCAAG AAAGCTGAAG
TACCGCTCTC TTTTCTAAGT CTAGTCTCTC TAGCTGTTGC GTTGCGGCTC TGTTCACTGC AAAAGCTTTG CTTCTCTTCC CGAAAGATTG TTTCTGACTTC

120 140 160 180 200
AACTCTCCGT TCTCTGCGAC GCGCTAGTGC CTCTCATCAT CTCTCTTCC ACCGAAAAAC TGTTGAGTGT CTGTAGTCTC AGGTCTTTCT TTCTCTCTCT
TTGAGAGGCA AGAGAGCGCT GCGCTACAGC GAGAGTAGTA GAAGAGAAGG TGGCTTTTGT ACAAGCTCAA GACATCGAGG TCCAGAAAGA AAGAGAGAGA

220 240 260 280 300
AACTTCCCTC TCTATAGATT TCTCATAACT CATCGAAGA ATCTTGTCTA GATCCAGACA AAAAATTTTA AAGAGTTTTT AGATGTATAT CTGATACATA
TTGAAGGGAG AGATATCTAA AGAGATTGGA GTAGCTTCTC TAGAACAGAT CTAGGTCTGT TTTTGTAAAT TTCTCAAAAA TCTACATATA GACTATGTAT

320 340 360 380 400
GGAGTTTACT GTATCAATCT TTATAGGACC ACTAATCTAT TATATAATTA AAATAGTTGT TAGAAACATT AATCATGACC ATAAATGACA TATATAAGT
CCTCAAATGA CATAGTTAGA AATATCTTGG TGATTGTAA ATATATTAAT TTATCAACA ATCTTTGTAA TTAGTACTGG TATTACTTGT ATATATTTCA

420 440 460 480 500
GTATAGTAAA ACTCTGTATT TAGATAAATT AAGGTATCTA ACTACGGTAA TATTCAAAAA GATGTAAATC TGGATATGCA TATATGTATA TTATTAGTAT
CATATCATTT TGAGACATAA ATCTATTTAA TTCCATAGAT TGATGTCATT ATAAGTTTTT CTACATTTAG ACCTATAGCT ATATACATAT AATAATCATA

520 540 560 580 600
ATAAATACAT GCTCTATAGT AGGTATTTGT GTCAACCATG TATAAATCTA TGTATATAGA TATTGTGGTA TGATATGTTT AAGCCGTCAA TGTCATATTT
TATTTATGTA CGAGATATCA TCCATAAACA CAGTTGGTAC ATATTTAGAT ACATATATCT ATAAACCCAT ACATACAAAA TTCCGGCAGT ACAGTATAAA

620 640 660 680 700
ATATAGAAAT ATGTGCGTAC CATAACATGA GGAAGTATCT ATATGTGTGT ATGTATAAAG CTTTCCCTTT GAAGAAGTAA TCTAAAAATA ATATATATAT
TATATCTTTA TACACCCATG GTATTGTACT CTTTCTATGA TATACACACC TACATATTTT GAAAGGGAAA CTTCTTCTAT AGATTTTTAT TATATATATA

720 740 760 780 800
ATATATGTAT ATGTATAGAT ATGTGTGAAT CTTTATATGT GTTGGGAAAA GTCAATTTAGA GAGATATTAT TGATATTAGG GATCTAAAT GACTTATCTG
TATATACATA TACATATCTA TACAACCTTA GAAATATCA CAACCTTTTT CAGTAAATCT CTCTATAATA ACTATAATCC CTAGATTTTA CTGAATAGCA

820 840 860 880 900
ATTACAGAGA TACGATTTTG GATTTTGTAC CCACTAGTTA TCAAGTCAGT TCCTATCTTC GGGGACATAC ACACCTTCAC AGATAATGT GTATATATGT
TAATGTCTCT ATGCTAAAAC CTA AAAAATG GGTGATCAAT AGTCGATCA AGGATAGAAG CCCCTGTATG TGTGAAAGTG TCTATTAACA CATATATACA

920 940 960 980 1000
AACTGAAAC GATAGTTTGA ACATGAATA ATGTACATGT TTGGGATTA ATGTGTTTTG TGGATTGGT TTGCATCTTT TGATTTTGA TTTTGTATA
TTGACTTTTG CTATCACAAT TGTACTTTAT TACATGTACA AACCCTAATT TACACAAAA ACCTAAACCA AAGCTAGAAA ACTAAATCT AAAACCATAT

1020 1040 1060 1080 1100
TTGTCGGTGT TTACATATGC ACATTGTTAA TATCAACAGT ATAGTTGTTT ATAAATAGTT ATTTATTGGA ATGTGTTTAT ATTATGAAG ATGAAGGAAG
AAGACCCACA AATGTATACG TGTAAACATT ATAGTTGTCA TATCAACAAA TATTATTCAA TAAATAACCT TACACAAATA TAATACTTCG TACTTCTCTC

1120 1140 1160 1180 1200
TCTTAGAGAG GCATAAATTG CAGTCAAGA ACTTGGAGAA GCTTGATCAG CCATCTCTTG AGTTACAGGT TAGCTACATT CTCGAAACGA CCACACATTT
AGGATCTCTC CGTATTGAAC GTCACTTTCT TGAACCTCTT CGAAGTAGTC GGTAGAGAAC TCAATGTCCA ATCGATGTA GAGCTTTGCT GGTGTGTAAA

1220 1240 1260 1280 1300
TCTTTCCGA TTCTGTAAAC TTGCAAAATC GAGTATTACT CCGTTGAATT ACCAATATGT TTTAGATTGT TGTAATTATT GACCAAGAAAT CTCTTAAATC
AGAAAGGGCT AAAGACATTT AAGCTTTTGT CTCATAATGA GGCAACTTAA TGGTTATACA AAATCTAACA ACATAATTA CTGGTTCTTA GAGATTTTGT

1320 1340 1360 1380 1400
TTTGTATTAA TAGGTACAAA ACTTTATATT ATTCATATG ATTAATTAGA CTGCATCCAT GTAGTAGTCA TGTAGAGTAG TCTGTGTAG AGAGTTGAGC
AAACATAATT ATCCATGTTT TGAAATATAA TAACGTATAC TAATTAATCT GAGCTAGGTA CATCATCAGT ACATCTCATC AGGACACATC TCTCAACTCG

1420 1440 1460 1480 1500
TTTAGATCAT TATGATATG ATTAAGAGCT TAAATCAATG TTTTATCTG TTAGCTGGTT GAGAACAGTG ATCACGCCCG AATGAGTAAA GAAATTCGGG
AAATCTAGTA ATACCTATAC TAATCTCGA ATTTAGTTAC AAAATAAGAC AATCGACCAA CTCTTGTAC TAGTGGGGG TTACTCATTT CTTTAACGCC

1520 1540 1560 1580 1600
ACAAGAGCCA CCGACTAAGG TACGTTATAT ATGTATATTC TATGACTTTT GAACCTAATA TCAATTTCTA ACTAATTTTT TTTTGTATCA ACCACTATCA
TGTTCTCGGT GGCTAGTTC ATGCAATATA TACATATAAG ATACTGAAAA CTGTATTGAT AGTAAAAAGT TGATTAATAA AAAAATAGT TGGTGATAGT

1620 1640 1660 1680 1700
TTTTCTAACT GTGTGTTTAC ATGATCATAT ATAGGCAAAAT GAGAGGAGAG GAACTTCAAG GACTTGACAT TGAAGAGCCT CAGCAGCTAG AGAAGGCCCT
AAAAGATTGA CACACAAATG TACTAGTATA TATCGGTTTA CTCTCTCTC CTGGAAGTTT CTGAACTGTA ACTTCTCGAA GCTCTCGATC TCTTCCGGGA

1720 1740 1760 1780 1800
TGAACTGGT TTGACGCGTG TGATTGAAAC AAAGGTTGTT AAGAAAAATA CTTGATACCA TGTATAAGTT TCTCTAAGCT TACGAGTATG CAATTTACTA
ACTTTGACCA AACTCGGCAC ACTAATTTG TTCTCAACAA TTCTTTTAAAT GAACTATGGT ACATATTCAA AGAGATTGGA ATGCTCATAC GTTAAATGAT

1820 1840 1860 1880 1900
ATACGAGATG TGTGTGAGA GTGCAAGAT TATGAGTGAG ATCAGCGAAC TFCAGAAAA GGTAAATAAT AACCAGAAATA ACGTTTATTC TTTACTTGAT
TATGCTCTAC ACAAGCTCT CACTGTCTTA ATACTCCTC TAGTCGTTG AAGTCTTTT CCATTAATTA TTGTTTATTA TGCAATAAG AAATAGACTA

1920 1940 1960 1980 2000
GATTTCAATA TTAATTTTGG CAGTTTCAAG ATCCAAATTT TTTACTTCT TCTCTTTTTT TTTGGTGTTC AGGGAAATGCA ATTTAGTGAT GAGAACAGC
CTAAAGTTAT AATTAAACC GTCAAAGTTC TAGTTTITTA AAGTAGAAGA AGAGAAAAA AAACCACAG TCCCTTGTTC TAACCTACCTA CTCTTGTTCG

2020 2040 2060 2080 2100
GGTTGAGGCA GCAAGTATGT GTCTTACCT CTCTGTGTAT AACAATCCC TTTCTTTTGT CTACCATTA CGTACACACC CCTAAATTTA ATCCCGAGT
CCAACCTCGG GGTTCATACA CAGAATGGGA TGTGTTAGGG AAAGAAAAA GATGGTAATT GCATGTGTGG GGTATTAAAT TAGGGGTCAA

2120
GTCTACAACA CATATGTTG ATCATACTGT GAGA
CAGATGTTGT GTATACAAAC TAGTATGACA CTCT

SEQ ID NO:46

Arabidopsis AGL24 genomic sequence

5
10
15
20
25
30
35
40
45
50
55
60
65

AGACTTACAA TAACTTCATC AAGCAACTCA TACACGAGCA CAAAGTTTTT CCTGAATGAA TCTTCATTCA GAACACCAAG ATAATCCTTA ATAACACGAG
TCTGAATGTT ATTGAAGTAG TTCGTTGAGT ATGTGCTCGT GTTTCAAAAA GGACTTACTT AGAAGTAAGT CTGTGGTTTC TATTAGGAAT TATTGTGCTC

CAATCCTTTG TAGAAGCTCC AAAACAAGAG AGGGTGACAC GTTAACTCTC GTTGTGCGAA CAAAATATAG ACCAACACCC TTGACATGGA AGTAGTTCAC
GTTAGGAAAC ATCTTCGAGG TTTTGTCTC TCACACTGTG CAATTGAGAG CAACAGCGTT GTTTTATATC TGGTTGTGG AACTGTACCT TCATCAAGTG

GCCATCGACA TTCTATAAGC AAAAAAATA AGTTAGATGA AATCATTACA GCTCACAACC AAACAGAAAG TATAATACCT ACAAAGATAG GTGGCGCCTC
CGTAGCTGT AAGATATCG TGTTTTAT TCATCTACT TTAGTAATGT CGAGTGTGG TTTGTCTTC ATATTATGA TGTTCATC CACCGCGGAG

TGCATTGCCA TCCTCCTCC AGAAGTTGAC TTACGGAAG AATGTCCTG TACTTCCTTT GGGTACCTCA GCCCGGTCTG TAGCAATAAA ACGTTACACA
ACGTAAACGT AGGAGGAAGG TCTTGAAGT AAATGCTTC TTACAGAGAC ATGAAGGAAA CCAATGGAGT CGGGCCAGAC ATCGTTATTT TGCAATGTGT

TCTTGAAACT TGTATTGGAT CCAACCAAT CGTATAATCT CAAAACAAAT AGCTTCTTTC TACTACATTA CATACAGATA CTCTGCCAA ACTAATTGAA
AGAAGTTGA ACATAACCTA GGTGGTTTA GCATATTAGA GTTTGTGTTA TCGAAAGAAG ATGATGTAAT GTATGTCTAT GAGACGGGTT TGATTAACCT

TAGTTTGTCT ATATTGTAT AATCTGATT GGAATTCAG CTCAACATAA TTTGTCATCG GATAAGAAAT GTTGTAGAT CAAACAGATC AATGAGCTTA
ATCAAAACGA TATAACATG TTAGACTAAA CCTTAAAGT GAGTTGTAAT AAACAGTAGC CTATTCTTTA CAACCATCTA GTTTGTCTAG TTAAGTGAAT

GAGAAGATTT CAATGGAAAA TTCTCATGAA ACAGTGACAT AAGACTCGAC TCTGAAGAGA AAAAGCAAAA CAGGAAGAA CAGAGAGGAT CAGATCGAGA
CTCTTCTAAA GTTACCTTTT AAGAGTACTT TGTCACTGTA TTCTGAGCTG AGACTTCTCT TTTTCGTTTT GTCTCTCTTA GTCTAGCTCT

AAGAGAGCTT ACAGTCACGG AAGACGATGT TATCTCCTCG CTGAGATAAG ACGAAGAATT GGGAGATCAT CATCGTTCCT TATAGCGGTG GATTCCGACT
TTCTCTCGAA TGTCACTGCC TTCTGCTACA ATAGAGGAGC GACTCTATTC TGCTTCTTAA CCTCTAGTA GTAGCAAGGA ATATCGCCAC CTAAGGCTGA

GTTCACCGC GAGTTTGGT AAGTCTACTG ATCGCCGATC GGTCTCGTCT TTTTGTGTGT CTGGTGGTGA GGTGGTTCAC GTTTTACCAT TTGCGGTGCT
CAAAAGTGCG CTCAAAACCA TTCAGATGAC TAGCGGCTAG CCAGAGCAGA AAAACACACA GACCACCACT CCACCAAGTG CAAAATGTTA AACGGCAGCA

TATCGTGAAG CTCTTTCATG AGACGGAGGG TTCTGTGTTT TTGTGAATTA TGATTTCCTG TTCTTATATG GGCCTATTTT TAAGACATCA ATATGGCCCA
ATAGCACTTC GAAGAAGTAC TCTGCTTCCC AAGACACAAA AACACTTAAT ACTAAGAAAC AAGAATATAC CCGGATAAAA ATTCTGTAGT TATACCGGGT

AATTTGGAAC TTGTATGAG TTTAAGGAAA TAAGTAGTAA GTACTATAAA TGATGGTTCG ATCTCGGAGG AGAAAAAATA AAACATTGTT TACGAGGAAAG
TTAAGCTTG AACATACTC AAATTCCTTT ATTCATCATT CATGATATTT ACTACCAAGC TAGAGCTCC TCTTTTTTTT TTTGTAAACA ATGCTCCTTC

CAAAATGTGA GTTGATATAA AGGGTACAAC ACATAATTTA TTTTGGGAAG TCAAAACTTT GAGGATTAAG CTGACAACGA AGGTAGTGA AGACTTTCGG
GTTTACACT CAACTATATT TCCCATGTTG TGTATTAAAT AAAAACTTC AGTTTTGAAA TCCTTAATTC GACTGTTGCT TCCAATCACT TCTGAAAGCC

GATCGAGCAA TCGGAGATA TACATGAGCC TAGAGGGCTG ACAAGATGAC CAAGCATTC AAATGAAAGG CTTAAGATTT TTCTTTTCT AAACCTCAAGT
CTAGCTCGTT AGCCCTCTAT ATGTACTCGG ATCTCCCGAC TGTCTACTG GTTCGTAAG TTTACTTTCC GAATTTCTAA AAGAAAAAGA TTTGAGTTCA

AAGAAACACA AGATATATGA AAGGTAACA AGGGTCAACA ACAAGTCTAA GCTTTTAAAG CGTGTAGAT GATTCTTCTT GAACACTATT ACAATTACTG
TTCITTTGTT TCTATATACT TTCCATTTG TCCCAATGTT TGTTCAGATT CGAAAAATTT GCACAATCTA CTAAGAAGAA CTTGTGATAA TGTTAATGAC

TTTGTGTTCA CATTATATG ACCTTGGGAG TCTTCTAGCT CGTCCCAAT ATATTCTCA CATATTACTA TAAGATCCTA AAGACCAATA ACATTGATCT
AAATCAAGT GTAAATATAC TGGAAACCTC AGAAGATCGA GCAGGGTTTA TATAAAGTT GTATAATGAT ATTCTAGGAT TTCTGGTTAT TGTAACTAGA

ACACCAAAAA CTCTCACTTT CTGATTTTGC ACTGCTTTT TTTCTCTCCA TAAACAAAAC CAAAGGCTTA CAATACTAAA TCTGTCTCAC ATTCTTAGTG
TGTGTTTTT GAGAGTGAAA GACTAAAACG TAGAGGAAAA AAAGGAGGGT ATTTGTTTTT GTTTCCGAAT GTTATGATTT AGACAGAGTG TAAGAATCAC

CTTATTTGTT TTAGTCATAA AGAAGTTAAT CTTATACAGA TTGAAGTCTT AAAGTCATCT ATATTACTTT TCACATGAT CATTATGAGA TGGTACGTTT

5
10
15
20
25
30
35
40
45
50
55
60
65

```
GAATAAACAA AATCAGTATT TCTTGAATTA GAATATGTCT AACTTCAGAA TTTCAGTAGA TATAATGAAA AGTGACATA GTAATACTCT ACCATGCAAA

-1281 -1261 -1241 -1221 -1201
CCCACGAATT TTATCAGTTT AGTTTAAATTT TCAGTTGTAC TTTGGGAGAA AAAATTTACA AGATACTTGT CGGCCATGAT ATCACCTTAG AGTTACCGGA
GGGTGCTTAA AATAGTCAAA TCAAATTTAA AGTCAACATG AAACCTCTTT TTTTAAATGT TCTATGAACA GCCGTACTA TAGTGGGATC TCAATGGCCT

-1181 -1161 -1141 -1121 -1101
GTCCGTGTAT ATATCATTTT TAATTAGGGT TAAACTTAA AAGGGTATAA ATGGCTGATC AAACCCAAAA ATAAAGATA ATGATGACGG TGGGAGACGA
CAGGCCACTA TATAGTAAAG ATTAATCCCA ATTTTGAATT TTCCCATATT TACCGACTAG TTTGGGTTTT TATTTTCTAT TACTACTGCC ACCCTCTGCT

-1081 -1061 -1041 -1021 -1001
GTGATCTTAT CAGGTGTCGC ATCTAGCATA TATAGGTGAA AGACTATAAA AAAGACATGA AATATTTAAT AGACACAAC TTTGTAATAA ACCAAAACCA
CACTAGAATA GTCCACAGCG TAGATCGTAT ATATCCACTT TCTGATATTT TTTCTGTACT TTATAAATTA TCTGTGTGA AAACATTATT TGGTTTTGGT

-981 -961 -941 -921 -901
AAAAGGTAGA TGAAGTATG AACAGCATCT TCTAATTACG AATAAAAAAA GTAACCAAAAC TTTCTTTCCA TTAGAATTGG TACGTAGTTC CTGTGTATT
TTTTCCATCT ACTTGACTAC TTGTCGTAGA AGATTAATGC TTATTTTTTT CATTTGGTTTG AAAGAAAGGT AATCTTAACC ATGCATCAAG GAACACATAA

-881 -861 -841 -821 -801
GTGATTTCTT TCAATTTCCA ATTATGTTTT TTTATTTTAT CATGTACAT TTTTGATAGT GGGTAAC TTTTATCTTT TATTTGACCT AGCCATATAT
CACTAAAGAA AGTAAAGGT TAATACAAA AAATAAATA GTACAATGTA AAAACTATCA CCCATTGAAA ACATAGTAAA ATAACTGGA TCGGTATATA

-781 -761 -741 -721 -701
AAATCTATTA ACTTATACGG AGTAGTATT CACGTCATTT ATTTTATTT TGTTTTAGA TGGGAAGTTA TTCAAACTA GACTAAAACA GTAAACTAG
TTTAGATAAT TGAATATGCC TCATCATAAA GTGCAGTAAA TAAAAATAA ACAAATCT ACCCTTCAAT AAGTTTGTAT CTGATTTTGT CATTTTGTATC

-681 -661 -641 -621 -601
GAAACCCGCT ACTGAATAAA GTTACAATTC CACATTATTC CATGACAGAC TAATTGAATT AGAAGGTAG GTAAATTTAT AAATCATAC TGTAGCAGTC
CTTTGGCGA TGACTTATTT CAATGTTAAG GTGTAATAAG GTACTGTCTG ATTAACTTAA TCTTCCAATC CATTTAATAA TTTAGTATTG ACATGTCAG

-581 -561 -541 -521 -501
TCTTCGTCTG GCAGCTCAGT CAGACAAAAC ACAAGTGTG TTTTGTGTT ATTTTAAATG ATTATAGTTT GGGAAAAAGA CATAATCAA AGGATACAA
AGAAGCAGAC CGTCGAGTCA GTCTGTTTTG TGTTCACAC AAATACACAA TAAAAATTAC TAATATCAA CCCTTTTCTC GTATTAGTTT TCCCTATGTT

-481 -461 -441 -421 -401
AACATATGGC CCATTGATAA GTATAGATCA CTGTTTAGCT AAAAAAGCA GACTCTTTTT TCCAATCTTG AACACAAACA CAGTCACCAT CTCTCTCTCT
TTGTATACCG GGTAACATT CATATCTAGT GACAAATCGA TTTTTTCGT CTGAGAAAAA AGGTTAGAAC TTGTGTTTTG GTCACTGGTA GAGAGAGAGA

-381 -361 -341 -321 -301
CTTTCTCTCT CACTCACACA TTAGGGAGTA AACAGCTACC AGAAAAACCT TTTTATCTT CTCACAAAT TAATAAGTGT GGTGCTGAGA TTGAATAACG
GAAAGAGAGA GTGAGTGTGT AATCCCTCAT TTGTCGATGG TCTTTTGGGA AAAAATAGAA GAGTGTTTAA ATTATTTTCA CCACGACTCT AACTTATTGC

-281 -261 -241 -221 -201
TAATCCAAGA TCCTCAACT CACAGAAAGG TAAAGCTGTG GAATCTGTGT TCTTCTTCT TAAGCAAAGT GTTTGATGAA TTCATCTAGT CCTGTCCATT
ATTAGTITCT AGGAGGTGTA GTGCTTTTCC ATTTTCGACA CTTAGACACA AGAAAGAAGA ATTCGTTTCA CAAACTACTT AAGTAGATCA GGACAGGTAA

-181 -161 -141 -121 -101
CTTTGCTTCT TCATGTTTGA TGGATCTGAT CTCTCTTCT CTCTCTCTCT AGCCATTAGG GTTCTCTAAG AATATTATAT AAACCTCTCT TAGCTAACAC
GAAAACGAAG AGTACCAAT ACCTAGACTA GAGAGAAAGA GAGAGAGAGA TCGGTAATCC CAAAGGATTC TTATAATATA TTTGAGAGAA ATCGATTGTG

-81 -61 -41 -21 -1
CGTTCCAATT GGTTCCTTTC TTTGTTCTTG GTCTAAAATC TAAATGTTGT TATGGGTATA GGCAGATTCA AGAACAGTAG TGAAGGAGAG ATCTGGTAAA
GCAAGGTTAA CCAAAGAAAG AAACAAGAAC CAGATTTTAG ATTTACCACA ATACCATAT COGTCTAAGT TCTTGTCTAT ACTTCTCTCT TAGACCATTT

20 40 60 80 100
ATGCGCAGAG AGAAGATAAG GATAAAGAAG ATTGATAACA TAACAGCGAG ACAAGTTACT TTCTCAAAGA GAAGAAGAGG AATCTTCAAG AAAGCCGATG
TACCGCTCTC TCTTCTATTC CTATTCTTTC TAACATTGTG ATTTGCGCTC TGTTCATGA AAGAGTTTCT CTTCTCTCC TTAGAAGTTC TTTGCGCTAC

120 140 160 180 200
AACTTTTCACT TCTTTGCGAT GCTGATGTT CTCTCTTCT CTCTCTGCCC ACCGGAAGC TCTTCGAGIT CTCCAGCTCA AGGTATATTC TATCTTTTGT
TTGAAAGTCA AGAAACGCTA CGACTACAAC GAGAGTAGTA GAAGAGACGG TGGCCTTTG AGAAGCTCAA GAGGTGAGT TCCATATAAG ATAGAAAAAC

220 240 260 280 300
TTAGTAGTGT TCTTATTTTT TTCAATCCAT GTTTGTGTT TTGAGAATAT GGTGGATAA ATATATTAAG ATATGTATTT AAATGAGATT TTTATTTTCT
AATCATCAAC AGAATAAAAA AAGTTAGGTA CAAACACAAA AACTCTTATA CCAACCTATT TATATAATTC TATACATAAA TTTACTCTAA AAATAAAGA

320 340 360 380 400
CGTTTACTCT CTAAAGTTAA TTATCAGTAG GCTCGAGAT CTCATGTACG GCATAATTG ATGACCTAAA TTATTACT TTAAGTATA GGATTGATGT
```


5

10

15

20

25

30

35

40

45

50

55

60

65

GCAAATGAGA GATTTC AATT AATAGTCATC CGAGCCTCTA GAGTACATGC CGTATTAAAC TACTGGATT TTAATAATGA AATTCATAT CCTAACTACA

420 440 460 480 500

TTTATTACTT TTATGTATA CACATCATGT ATTAAATCC GTTTAACATA ATATGGGTTT TTAACGTGTA ATTTTTCAT CATTTCATT TAGACTCATG

AAATAATGAA AATACATATT GTGTAGTACA TAAATTAAGG CAAATTGTAT TATACCCAAA AATTGCACAT TAAAAAGTTA GTAAAAAGTAA ATCTGAGTAC

520 540 560 580 600

GTTAAGATT CTGTACTGGG AAATAAGAGA GCAGAAATAT ATAGTGTGAT TTTTGTAAAT TAGGAAAGCA TATGTATATA TGGATACATA GTACTTACCA

CAATTCTAAA GACATGACCC TTTATTCTCT CGTCTTATAA TATCACACTA AAAACAATTA ATCCTTTCTG ATACATATAT ACCTATGTAT CATGAATGGT

620 640 660 680 700

CAATTAGAA GAATTTCTTT TCCCTTTTCT CATTGACTT TGTGTATTAC AAAAGTCTTT GACACTGTCA CTGTGTATGA TTGGGGATTA ATCTTAAACC

GTTAATCTTA CTTAAAGAAA AGGGAAAAAA GTAAACTGAA ACACATAATG TTTTCAGAAA CTGTGACAGT GAACCATACT AACCCTTAAT TAAGAATTGG

720 740 760 780 800

ACTCGTTTAT TTTATCTGGG GAAGCATTAC CATAATGGG AAACGAGTCA TCTGTCTGTA TCGTGATGGC TACTTCTGAT TACTTTTCTT TTATTATAAC

TGAGCAAAAT AAATAGAACC CTTCGTAATG GTATTAAACC TTTGCTCAGT AGACAGACAT AGCACTACCG ATGAAGACTA ATGAAAAGAA AATAATATTG

820 840 860 880 900

CAAAAAGGCT TCTAATGTAC TTAATTAAAT TTACAAATGT AATATGGACG AAGGAAATGT TTATAAGAAA GATGGATTGT TTGTTGAAAC GTGTAGAATG

GTTTTTCGGA AGATTACATG AATTAATTAA AATGTATTACA TTATACCTGC TTCTTTTACA AATATTCTTT CTACCTAACA AACAACTTTG CACATCTTAC

920 940 960 980 1000

AGAGACATAT TGGGAAGGTA TAGTCTTCTC GCAAGTAACA TCAACAAATT GATGGATCCA CCTTCTACTC ATCTCCGGGT ATTTTCGATA TCACCTACTC

TCTCTGTATA ACCCTTCCAT ATCAGAAGTA CGTTCATTGT AGTTGTATTAA CTACCTAGGT GGAAGATGAG TAGAGGCCCA TAAAGCTAT AGTGAATGAG

1020 1040 1060 1080 1100

TTTTTTTTTT TTGTGGATT TAAACTCTCT AAGTCTTTTA CCAACCCTTT CTCTTTTAT CAAACCCTTC TCTCTATAAT ATTATCCGAT GTTCACCTTG

AAAAAAGAAA AACACCTAAA ATTTGAGAGA CGAGAAAAAT GGTTTGGGAA GAGAAAAATA GTTTGGGAAG AGAGATATTA TAATAGGCTA CAAGTGAAC

1120 1140 1160 1180 1200

TTACACGTGT TTGTATAAT TTTAGCTGT AGTCTAAAT ATAGAAACAT TGAGTGGCAT ATAATCATT TCTTGAAGC ATCTAATTAA TTGGTTTATC

AATGTGCACA AACAAATTA AAAATCGACA TTCAGATTTA TATCTTTGTA ACTCACCGTA TATTAGTAAT TAGAACTTCG TAGATTAAAT AACCAAAATG

1220 1240 1260 1280 1300

ATATTAAATG CAGAACTCTG AAACGTGTAT CTTTGCATCT AGCAGCTTGA GAAATGTAAC CTCTCCAGAC TAAGTAAGGA AGTCGAAGAC AAAACCAAGC

TATAATTATC GTCTTAGGAC TTTGACAAC TAAACGTAGA TCGTCGAAC CTTAACATTG GAGAGGTCTG ATTCATTCTC TCAGCTTCTG TTTTGGTTCTG

1320 1340 1360 1380 1400

AGCTACGGTA TGGCTCCATT GATATGTTAT GCAGATAAAC CTATTTTTCAT ATAGGCTATA GCTGTAAGAG ATCATCTATT TCATGTGTGT GGTTTTTTTT

TCGATGCCAT ACCGAGGTAA CTATACAATA CGTCTATTTG GATAAAAGTA TATCCGATAT CGACATTCTC TAGTAGATAA AGTACACACA CCAAAAAA

1420 1440 1460 1480 1500

TTTATGTTTT TTCAATGATG TGTGCATGCT ATTTTAGGT TTTAGAATCT ATTTCAATGA AATTGAAGAT ATTTCAATTC ACGTGAAGT TCGTCAAGTT

AAATACAAAA AAGTTACTAC ACACGTACGA TAAAAATCCA AAATCTTAGA TAAAGTACCT TTAACCTCTA TAAAGTAAAG TGCACATTCA AGCAGTTCAA

1520 1540 1560 1580 1600

GTGGCGTGTG TCTTGGAAAT TGATGTTTTT TTGTAGATT TTAAGAGCTA CTTCTAAAT TTACAAGAGT TTGTGAATTT TCAATTATGG CCCATTATTC

CACCGCACAC AGAACCTTAA ACTACAAAAC AAACATCTAA AATTCTCGAT GAAGATTTTA AATGTTCTCA AAACATTAAA AGTTAATACC GGGTAATAAG

1620 1640 1660 1680 1700

TCATTAAATC ATTAATAAAA TTATATACAT TACTATCTAT ATCTAGCATA GGTAGTTTTT TTTTCTTTT TCTTTGGTAG ACCTACTGAA CAAATATCTG

AGTAATTAAG TAATTTTTTT AATATATGTA ATGATAGATA TAGATCGTAT CCATCAAAAA AAAAAGAAAA AGAAACCATC TGGATGACTT GTTTATAGAC

1720 1740 1760 1780 1800

ATATATCACT GACTGGATAA ATATCTATAG AGATATTTTT GATAGAAATG AGTGTTAATT TAACGTAAAA CAGGAAACTG AGAGGAGAGG ATCTTGATGG

TATATAGTGA CTGACCTATT TATAGATATC TCTATAAAAA CTATCTTTAC TCACAATTAA ATTGCATTTT GTCTTTTGAC TCTCTCTCC TAGAACTACC

1820 1840 1860 1880 1900

ATTGAACCTA GAAGAGTTGC AGCGGCTGGA GAACTACTT GAATCCGGAC TTAGCCGTGT GTCTGAAAAA AAGGTTTACT ACTATACATA AACTAATAGC

TAACCTGAAT CTTCTCAACG TCGCCGACCT CTTTGATGAA CTTAGCCCTG AATCGGCACA CAGACTTTTT TTCCAAATGA TGATATGTAT TTGATTATCG

1920 1940 1960 1980 2000

ATGCATATTT TCCTTAACGT GGCATATAAA TAATAAGCTG TACATATATA AAAGTTTGAC TTTGTTGTTG TTAATTGTAA ATAGGGCGAG TGTGTGATGA

TACGTATAAA AGGAATTGCA CCGTATATTT ATTATTGAC ATGTATATAT TTTCAAACCTG AAACAACAAC AATAACCAT TATCCCGCTC ACACACTACT

2020 2040 2060 2080 2100

GCCAAATTTT CTCACPTGAG AAACGGTTTA GTAGTTAGTA CATACAATTC GTATAACTAA TGGATCATAA GCCTATCTAT AGCTAGTGAC TTTCTTAATA

CGGTTTAAAA GAGTGAAC TC TTTGCCCAAT CATCAATCAT GTATGTTAAG CATATTGATT ACCTAGTATT CGGATAGATA TCGATCACTG AAAGAATTAT

2120 2140 2160 2180 2200

AGTGAAACAG GGATCGGAAT TGGTGGATGA GAATAAGAGA CTGAGGGATA AAGTACGGCT CTAAACCCCT ATAGATATCA TGGAAATAACC TTAATCTATT

TCATCTTGTC CCTAGCCCTA ACCACCTACT CTTATCTCTT GACTCCCTAT TTCATGCCGA GATTTGGGAA TATCTATAGT ACCTTATTGG AATTAGATAA

2220 2240 2260 2280 2300

TTTTTATGTA TAAGAAAAAT TGATGAGGGA ACGTATATTA TATATCGGCA GCTAGAGACG TTGGAAGGGG CAAACTGAC GACGCTTAA GAGGCTTTGG

AAAAATACAT ATTCTTTTAT ACTACTCCCT TGACATATAAT ATATAGCCGT CGATCTCTGC AACCTTTCCC GTTTTGACTG CTGGGAATTT CTCGGAATCC

2320 2340 2360 2380 2400

AGACAGATC GGTGACCACA AATGTGTCAA GCTACGACAG TGGAACTCCC CTTGAGGATG ACTCCGACAC TTCCCTGAAG CTTGGGTATA ATTTGTTTAA

TCTGTCTCAG CCATCGGTGT TTACACAGTT CGATGCTGTC ACCTTGAGGG GAATCCTTAC TGAGGCTGTG AAGGGACTTC GAACCATAT TAACAAATAT

2420 2440 2460 2480

CTGAACATAT TTCAAACTTT TTGTGTGACAT TTTGTATGTG GATGTTTACT AACTGTTGT TGGTTAGGCT TCCATCTGG GAA

GACTGTGATA AAGTTTGAAA AACAACTGTA AAACATACAC CTACAAATGA TTGACAAACA ACCAATCCGA AGGTAGAACC CTT

SEQ ID NO:47

Arabidopsis AGL27 genomic sequence

CAACCAGCAG CACCAGCTGC AATCAAATCC -2961 TTTACGGTTC TTTGAATGTT TAGCGCATTT -2941 CCTCCACCGG TATCTTGAAA AGATCAAAAG AAACCTATGA -2901

GTTGGTCTGC GTGGTCGACG TTAGTTTAGG AAATGCAAG AAACCTTCAA ATCGCGTAAA GGAGGTGGCC ATAGAACTTT TCTAGTTTTC TTTGATACT

-2881 -2861 -2841 -2821 -2801

AGAGAACTAT AACCAAGCAA ATCCACTATT TTCAAAGAGA ACTATAAGCA AGCAACGCAC TCTAACCAAG AAAGATTGAT ACTTTCAATC -2801

TCTCTTGATA TTGGTTCGTT TAGGTGATAA AAGTTTTCG ATACTCTCTT TGATATTGTT TCGTTGCTG AGATTGGTTC TTTCTAACTA TGAAAGTTAG

-2781 -2761 -2741 -2721 -2701

TTTGGTAAAG AATCAACGAC TCAATGTTTT TAAATGTTTT TTTTCTTTT TTGGTTTGTG TTAAGTTCTT TGCATTCTTT AATGATGTTT TTATTATACT -2701

AAACCATTTT TTAGTTGCTG AGTTACAAAA ATTTACAAAA AAAAGGAAAA AACCAAAATC AATTCGAAGA ACGTAAGAAA TTACTACAGA AATATATGA

-2681 -2661 -2641 -2621 -2601

ATCAAAATTT TGCAACTTTA CCAGCATCTG CAATGATGGG TATATTAGGA GCTGACGCAC ACACCGACCT TGCCGTCGCA GCCATCTCCG GTGGTCTAAA -2601

TAGTTTAAAA ACGTTGAAT GTTCGTAGAC GTTACTACCC ATATAATCCT CGACTGCGTG TGTGGCTGGA ACGGCAGCGT CCGTAGAGGC CACCAGATTT

-2581 -2561 -2541 -2521 -2501

ACGACGAAAG AACCAAGATA AAACGAAAGC ATCAAAACAA AAAATTACTA AAGAAAGAAA AAAAAAAGG TGCGCAGCTG TAGCAAAACG AAATCGGGTT -2501

TGCTGCTTTC TTGTGTTTAT TTGCTTTTCG TATGTTTGT TTTTAATGAT TTCTTTCTTT TTTTCTTCC ACCGCGTGCA ATCGTTTGGC TTTAGCCCAA

-2481 -2461 -2441 -2421 -2401

TTCCAGGAG AGAAGCGGAT AAGCGTAAAC CGGATATATA ACCAGCGGAG AATCCGGTTT GCTGCACAAT AGCCCGGAT AAGGCATCGT AGCATCCAGG -2401

AAGGTCCTC TCTTCGCTA TTCCGCTAGT GCCTATATTT TGGTCGCTC TTAGGCCAAA CGACGTGTTA TCGCGCCTA TTCCGTAGCA TCGTAGGTCC

-2381 -2361 -2341 -2321 -2301

CATAAGCACA ATGCCTTGTT CTTCAATCAG CGGATGAAAA CGTGTTTGGA TTCTCGTGT TTCTCGTGT CGGATTACCC AATCTCGGAT TCGCTCGGAT -2301

GTATTCGTGT TACGGAACAA GAACTTAGTC CGCTACTTTT GCACAAACCT AAGAGCGACA GCCTAAGTGG TTAGAGCGGC GCGCAGCCAA GCGTCTGGAAT -2301

-2281 -2261 -2241 -2221 -2201

GTTGGTGAAG CTGTAAGGTT TAAGCTGCTA CAACAGAGTT AAGTTGTTT GACAGCCATT TTTCTCGAAG CCTCGAACA GTTTTTTCTT -2201

CAACCATTC GACATTCCAA ATTCGACGAT GTTGTCTCAC TTCAACAAA TTGTCCGTAA TTGTAGCTGT AAGAAGCTTC GGAGCTTGTT CAAAAGAGAA

-2181 -2161 -2141 -2121 -2101

CTCTCTAATC GAGTTAGACT CTGACCCACA CGCTTGGGAT TTTAATAGAG AGCACGTGGT TATTATATCT CGGTCTTATC TTATGGTAA AGTATCTCAA -2101

GAGAGATTAG CTCAATCTGA GACTGGGTGT GCGAACCCCTA AAATATATCT TCGTGACCA ATAATATAGA GCCAGATAG AATACCATTC TCATAGAGTT

-2081 -2061 -2041 -2021 -2001

AGACTCAAA CACAAGGTAT TGTGAAATG TTAGAGGCAA TCTAACAAAT AATGTATAT TTGTTAGCT TAAGCTCATC ATAGAAATGG GCCTTTATGT -2001

TCTGAGTTTG GTGTGTCATA ACATCTTTAC AATCTCCGTT AGATTGTTAT TTACATATTA AACCAATCGA ATTCGATAG TATCTTTACC CGGAATATGA

-1981 -1961 -1941 -1921 -1901

CACCAAACTT ATTTCAACAA ATAACACAAG AGCCCAACAA ACAACGACTC CTTTCTCCAC CAGAAACAGC ACGCAAAAGC CAAGAGAGTT GCAAAAGACC -1901

GTGGTTTGGG TAAAGTGTG TATTTGTTC TCGGCTGTTT TGTGTCTGAG GAAAGAGGTG GTCTGTGTC TGCTGTTTCC GTTCTCTCAA CGTTTTCTGG

-1881 -1861 -1841 -1821 -1801

TATAAGATGA TAACAATCGA AAAGATGTAA ATTTTGAGAA AAATCAAAAT AAACAAGAAA GATTTCAATG TTTTCACTT TTTCTCAIT TCTACTTTGA -1801

ATATTCTACT ATTTGTAGCT TTCTACATTT TAAACTCTTT TTTAGTTTAA TTTGTTCTTT CTAAGTAAAC AAAAAGTGAA AAAGAGGTAA AGATGAACT

-1781 -1761 -1741 -1721 -1701

TTTTACATAC TCTATGGGCC AACCAATTC CAACCTAATG CTTGATAAAA AATGATTCGG TTTTACTATC TCAACAAAT GGGCTACAA CATCCAATTT -1701

AAAATGTATG AGATACCCGG TTGTTTAAAG GTTGGATTAC GAACATATTT TTACTAAGCC AAAATGATAG AGTTGTTTAA CCGGATGTT GTAGGTTAAA

-1681 -1661 -1641 -1621 -1601

CATGTAGTGA CTTGTTTGTG CTTTCTTTC ATCTCAACAA ATTTGGTCTG TTGTATTTAA GAAATGTGTA CAGCTTTTGA GACTGAATTT TACTTTATGG -1601

GTACATCACT GAACAAAAGC GGAAGAGGTG TAGAGTTGTT TAACCCAGCA AACATAAATT CTTTAACAAT GTCGAAAAAT CTGACTTAAA ATGAATATCC

-1581 -1561 -1541 -1521 -1501

CTTTATGCTC TCTTTTCCG TTTTGATTAA GGGTGAATAT GTAACTGTT GATACCATCT GATTTTCTTT ATTTTCTATT TTTCTGTGT GCAACTATAC -1501

GAAATACGAG AGAAAAAGC AAAACTAATT CCCACTTATA CATTTGACAA CTATGTTAGA CTAACAAAAA TAAAAAAATA AAAGAACACA CGTTGATATG

-1481 -1461 -1441 -1421 -1401

CATCTGAATT CAATTGACAT TTTAGCCAAA TAAAAAGAT TGGTCCACTT GGATGGCTGT AAAAAAGTTT AGTGGAGTA TTTATAGGGC TTGTTGGCAA -1401

GTAGACTTAA GTTAACTGTA AAATCGGTTT ATTTTCTTA ACCAGGTGAA CCTACCGACA TTTTTCAAA TCACCTTCAT AAATATCCCG AACACCGTT

-1381 -1361 -1341 -1321 -1301

TCTTCACCAA CGGCTAATAT GTTGATCTTT TTAATAATTA ACTTACCGTT CGACTGTCTT CTCAACGATT TGACAATTAG CCGTTAGATT AGTATTACTG -1301

AGAAGTGTTT GCCGATATTA CAACTAGAAA AATTTTAATT TGAATGGCAA GCTGACAGAA GAGTTGCTAA ACTGTTAATC GGCAATCTAA TCATATATGAC

-1281 -1261 -1241 -1221 -1201

ATTTATTATT AACAAACCA TTTCTTTTCT TATTTTGTAA TAAGCTAAT CAGGCCAATA AAAGGACAAA GTAGAGATGG GCTATTTCTT TTTTCTCTTT -1201

TAAATAATAA TTGTTTGGGT AAAGAAAAGA ATAAAAACTT ATTCGATTA GTCCCGTTAT TTTCCCTGTT CATCTCTACC CGATAAGAA AAAAAAGAAA

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

```
-1181      -1161      -1141      -1121      -1101
TTTTTTTTTC TTATGTAGTA GAGAAAAGCC TTTATTCCTA GAGCTATCAT TTACCACCCA TTAACCAGAA GCTGAGAAAT GAAGCAAGCC GAAACGAAAT
AAAAAAAAAG AATACATCAT CTCTTTTCGG AAATAAGAAT ATCGATAGTA AATGGTGGGT AATGGTCTTT CTTCGTTTCGG CTTTGGCTTAA

-1081      -1061      -1041      -1021      -1001
TGTAGTTTTC GACGGTGAAG TTATATCGGG CCTTTAATGG GCATGTGAAT AGAGTTGAGA GTCTTTTTCG CCCAAATAAT CGTTTAAAGG AGTATTGGCT
ACATCAAAAC CTGCGACTTT AATATAGCCC GGAAATTACC CGTACACTTA TCTCAACTCT CAGAAAAACG GGGTTTATTA GCAAAATCCC TCATACCCGA

-981      -961      -941      -921      -901
CGTTGGTTTA ATATTGGGCC GAAACGAGAT TGGGAAGAAG AACAATGTCT GTTTAATCCG GTTAGGGTCG TGGGCTGATT CTGGTTCACC TTTATAGCGT
GCAACCAAAAT TATAACCCGG CTTTGTCTTA ACCCTTCTTC TTGTTACAGC CAAATTAGGC CAATCCCAGC ACCCGACTAA GACCAAGTGG AAATATCGCA

-881      -861      -841      -821      -801
AAGCGAACAA ACATTGAAAA TGGGAAGGCC AAATTAGTTA CCATCCCTAA CTCAGTTTTC AGACGTAGTA TGAATGAGCC ACGGCAGAAC CTACGACCTA
TTGCGTTGTT TGTAACTTTT ACCCTTTCGG TTTAATCAAT GGTAGGGATT GAGTCAAAAC TCTGCATCAT ACTTACTCGG TGCCGTCTTG GATGCTGGAT

-781      -761      -741      -721      -701
ACTCGATAAA GTAATGGTTA CTCTTGAGAG CGGAAGAAGG CACAAGATT TTGATAAGGC TTTCTAGTTG GTGAAATGGT CAAAATCGCT CGGAGAGCCA
TGAGCTATTT CATTACCAAT GAGAACCTCT GCCTTCTTTC AACTATTCCG AAAGATCAAC CACTTTTACCA GTTTTAGCGA GCCTCTCGGT

-681      -661      -641      -621      -601
TCATAGGAGC GGGGAGGTGC TATCTGAATA TCCCAATGCA TCAAGACAAG ATCGATTTCG AAAACAAAGA AATTAAACAA ACATTTTAAA ATATGCTCTT
AGTATCTCTG CCCCTCCAGC ATAGACTTAT AGGGTTACGT AGTCTGTCTC TACCTAAGTC TTTTGTCTCT TTAATTTGTT TGTAAAAATT TATACGAGAA

-581      -561      -541      -521      -501
AGTTTATAGT AATATAAATG TTTCAATACC AATTATCTTA CACTGATAGT GGTCAAGTTA CTAATCACIT TTAATAAATT GGTGATAGTC AAACGTATTG
TCAAAATCTA TTATATTACA AAGATTATGG TTAATAGAAAT GTGACTATCA CCAGTTCATAT GATTAGTGAA AATTATTGTT CCACATACAG TTTGCATAAC

-481      -461      -441      -421      -401
AAAATTATCG ATTTAAAAAT ATTTGAATTC AAAACCAATT TAGTGAAAGT TTGCATTGTA GTTTTGATTA TCCGATCAAT CTTTAATATA ATTAGCTCAA
TTTTAATAGC TAAATTTTAA TAAACTTAAG TTTTGGTAAA ATCACTTTCA AAGCTAACAT CAAAACATAA AGGCTAGTTT TATATGAGTT

-381      -361      -341      -321      -301
TAATAACTGA AATCCTTGAAG TTAACCGTTA CCCGATTCTAT AAGCACTACT TTCCGATCAA AACCAATGAG ATAAATAAAC TTTTAAACCC TCCAATAAAA
ATTATTGACT TTAGGAACCT AATTGGCAAT GGGCTAAGTA TTCGTGATGA AAGGCTAGTT TTGGTTACTC TATTTTATTG AAAAATTTGG AGGTTTATTT

-281      -261      -241      -221      -201
AAGAGAAAAA CTTAAAAACC AATTTCGTGT CGGTGGGGGT GATGATCGGA CTCGGACCGG TCTAACCGAC TGGATTAAAA AGTCTTTAAC AACGACAAGC
TTCTCTTTTG GAATTTTTCG TTAAGACAAA GCCACCCCTA CTACTAGCCT GAGCCTGGCC AGATTGGCTG ACCTAAATTT TCAGAAATTT TTGCTGTTTC

-181      -161      -141      -121      -101
TTAAAAATTT GCCTCTTAGT GGCTTCAAAA CGCAATCGTT TCGCTTAATA CTATTATTTT CTCTATCTCG TTTAACCAAA AAAAAAACG AGTTGGAGGA
AATTTTAAAA CGGAGATCAA CCGAAGTTTT GCGTTTCAAA AGCGAATTAT GATAATAAAA GAGATAGAGC AAATTTGTTT TTTTCTTTTC TCAACCTCTC

-81      -61      -41      -21      -1
AAAAAAAAC CAAGAAAAAA GAATAAAAAA CAAAAGCAT TGAGCGTCTC CGGAGATTAG GATTAAATTA GGGCATAAAC CTTATCGGAG ATTTGAAGCC
TTTTTTTTTG GTTCTTTTTT CTATTTTTTT GTTTTTCGTA ACTCGCAGAG GCCTCTAATC CTAATTTAAT CCCGTATTGG GAATAGCCTC TAAACTTCGG

20      40      60      80      100
ATGGGAAGAA GAAAAATCGA GATCAACGCA ATCGAGAACA AAAGCAGTCG ACAAGTCACT TTCTCCAAAC GACGCAATAG TCTCATCGAC AAAGCTCGAC
TACCCTTCTT CTTTTTAGCT CTAGTTCGCT TAGCTCTTGT TTTCGTGAGC TGTTCACTGA AAGAGGTTTG CTGCGTTACC AGAGTAGCTG TTTCGAGCTG

120      140      160      180      200
AACTTTGATG TCTCTGTGAA TCCTCCGTCG CTGTGTCTGT CGTATCTGCC TCCGGAAAC TCTATGACTC TTCTCCCGGT GACGAGTAAG AAGATACTTT
TTGAAGCTA AGAGACACTT AGGAGGCAGC GACAACAGCA GCATAGACGG AGGCCTTTTG AGATACTGAG AAGGAGGCTA CTGCTCATT CTCTATGAAA

220      240      260      280      300
CCTTTTCTGG GTCTCACTCG AITTTTGTGC TTTTTCATCT TGTTTAATTA CTTTCTCCAT ATAGAAGCTT CAAATCTAGG GCTTTTGTAT TCCATCAAA
GGAAAGACCC CAGAGTAGGC TAAAAACACG AAAAATAGAA ACATAATTAAT GAAAGAGGTA TATCTTGCAA GTTTAGATCC CGAAAACTA AGGTATGTTA

320      340      360      380      400
CAACTGAGAT TTTCTCCTTG TTTTCTGTAT GAAGATAGCA GATGCGTAAG CTTTAACTTA ATTTAAGACT AAACATTTTG ATCGCAAGA TATGTTCTTG
GTTGACTCTA AAAAGAGAAC AAAAGACATA CTTCTATCGT TTAATTTGAT GAAATTTGAT TTTGTAAGCT TTTGTAAGCT TAGCGGTTCT TATCAAGAAC

420      440      460      480      500
ATGTTCTGTT CGTGTTTTTT TTTTCTGTGT TTTTPTTTTT TCATTTTAAA ATCATTTTTA TCTCTTTTTT TACCTTCATT TGTGACGAAA TTTAATATTG
TACAAGCAAA GCACAAAAAA AAAAGCACAA AAAAAAATTT AGTAAAAATTT TAGTAAAAAT AGAGAAAAAA ATGGAAAGTA ACACGTCTTT AAATTATAAC

520      540      560      580      600
CATGTTATTC AAGAAACTTT TCTACACGTG GTGATTCGTT CTTGATGTTC TTTAAGTAAT CTTTGTATTG CTAGTTCATC CTGTGTTTCA CTTTGAAGCT
GTACAATAAG TTCTTTGAAA AGATTGTGCAC CACTAAGCAA CTGATCTAAC TTAATTCATTA AAATTCATTA TAAACATAAC TACCAAGCTA GAAACTTCGA

620      640      660      680      700
TGGTTTTCCT ATATAAGAAA CAATATGTTT AGATTGTTCA AATTTTGAGA TTTGGTAATT ATATTCAATA TTGCAATTGA CTTCAAGTAG TTTTGTGAG
AGCAAAAAAG TATATCTCTT GTTATACAAA TCTAACAAGT TTAACACTCT AAACCATTTA TATAAGTTAT TACAATGCTA AACCTTACCT AAAACACTCT

720      740      760      780      800
AGATTATTTC GGGTTAGTGG TAACATTAAT CGAATATCTT TGGTTCAAAT TGGTTAACAC ATTGTACTTT TAACATGAAA ATGTGATCTT AAAATGTATT GTAGATCTAT
TCTAATAAAC CCCAATCACC AATTGTAATTA GCTTATAGAA TGAACAGTTA ACCAATTGTT ACCAATTGTT TAACATGAAA TACAATGCTA AATGACTTAC AATGACTCAC

820      840      860      880      900
TCTTTTGTAA TTCTCTTTAA GGAATAAGGT TTATCTAGTT GATTTTGATG GTTTATTGTA GTGCTGGGAT AAGTTTCCAC ATTGATACTC GCCACACATT
AGAAAACAAT AAGAGAAATT CCTTATCCCA AATAGATCAA CTAAAACTAC CAAATAACAT CACGACCCTA TCAAAAGTG TCAAAAGTG TAACATAGAG CGGTGTGTAA

920      940      960      980      1000
CTTCATTACT TAACTAAATT GATATCGATT TTAACCCITT TAATCGTAAT TTGTTGTGTG TTTATGACAC CATACAAGAT ACATTATGTC TTACTGAGTG
GAAGTAATGA ATGATTAAAC CTATAGCTAA AATTGGGAAA ATTAGCATTA AACAACACAC AAATACTGTG AATACTGTG GATATGTTCTA TGTATACAG AATGACTCAC

1020      1040      1060      1080      1100
ACTCTTTGTT GCTCTCTAAG ATGTTGTAGT TTGGAATTCT TTGCTAAAGA AACTCAAATC ATAACGTATT TTAAGTCTAT CATATATATG TCAGTGGCCT
TGAGAAACAA CGAGAGATTG TACAACATCA AACCTAAGAA TTAGATTCTT TTGAGTTTGA TATTGACTAA TATTGACTAA AATGACGATT GTATATATAC AGTCACCCGA

1120      1140      1160      1180      1200
AGTAGGTTCA TTAAGTAGAA ATCGGTGCGC AATTTTACTA ATTTGGAGAA ACCACTAGAC TACAACCAAA TGTTCATAGA CTTTAAATAG CTTCTGTTAT
TCATCCAAGT AATTCCTCTT TAGCCAGCGG TTAATAAGAT TAACCTCTTT TGGTATCTG TGGTATCTG ATGTTGGTTT ACAGTTTACT GAAATTTATCA GAAACATAA

1220      1240      1260      1280      1300
TTGTCGTGGA TATTTTAAAC CCCATGAAGT TTTGTATCTA GAAAAATCTC ATCCACTTCT CTTTGTAGAA ACTTTGAATG CGACTAAAAA TGAGTTTCTT
AACAGCACTT ATAAAAATTG GGGTACTTTG AAACATAGAT CTTTTGTAG TAGGTAGAAG GAAAACTCTA TGAACCTTAC GCTGATTTTC ACTCAAAAAA

1320      1340      1360      1380      1400
```

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100

TTTTCTAATA AAAAGATTAT	GACCTAAGAT CTGGATTCTA	AAAATCATCA TTTTAGTAGT	ATGGATAAGT TACCTATTCA	AGGAAATGGA TCCTTTACCT	AAGGTAACCT TTCCATTGAG	TTGTCACTAT AACAGTCATA	GTGTATATAT CACATATATA	ACAGCTCCTT TGTCGAGGAA	CTCATTTCTT GAGTAAAGGA
1420	1440	1460	1480	1500					
TGATGTTGAC ACTACAACCTG	TCCATAAATG AGGTATTCTAC	CTTGATCATG GAACATAGTA	AAAGCAAATT TTTCGTTTAA	TGTTAAATTT ACAATTTAAA	GTAACCAACA CATTTGGTTGT	AAATGCACAG TTTACGTGTC	ACTATAGACG TGATATCTGC	AAGTATTAGG TTCATAATCC	AACCGTATCT TTGGCATAGA
1520	1540	1560	1580	1600					
ATCTGTCTCC TAGACAGAGG	ATTTTACAAT TAAATGTGTA	AGTCAAGCTC TCAGTTCGAG	TAGTTGTAGC ATCAACATCG	TAGTTTCTTT ATCAAGAAAA	ATTTAGTTCT TAAATCAAGA	TATACCTTAA ATATGGAAAT	CAAAAGTGGA GTTTCACCGT	CTATGCAAA GATACGTTTC	TGTTTTTAGT ACAAAAATCA
1620	1640	1660	1680	1700					
TGAGATTAGT ACTCTAATCA	CGTCTTAGTC GCAGAATACG	GTCTTACTAA CAGAATGATT	TTGTTCAATT AACAAGTAAA	TTTCTTCTTT AAGAAGAAAA	TTGTGATTGA AACACTAACT	TGTAAATTA ACATTTTAAT	CTAAGTCACA GATTCACTGT	ACTTGAGATG TGAACCTCTC	TTACTAAAAA AATGATTTTT
1720	1740	1760	1780	1800					
GATAAGAACG CTATTCTTGC	TGTAATAACT ACATTATTGA	GAAGTGAATT CTTCACTTAA	TGAAGCCAGT ACTTCGGTCA	CTCTATTTCAT GAGATAAGTA	ATCATAGCAT TAGTATCGTA	TAATAGATCA ATTATCTAGT	TGGACAACAC ACCTGTTGTG	ATATATAGGA TATATATCTT	TTAGAGCTGT AATCTCGACA
1820	1840	1860	1880	1900					
CATGACCTTC GTACTGGAAG	CCGGAAATGC GGCCTTTACG	TAAATCAGTT ATTTAGTCAA	TCTTGGTTTA AGAACCAGAT	TCCTTTTGGG AGGAAAAACC	AGTATCATGA TCATAGTAA	TATCATTTAG ATAGTAAATC	CCAAAGGTTT GGTTTCCAAA	TTGGTTTCAG AACCAGAGTC	TATTCGAGTT ATAAGGCTAA
1920	1940	1960	1980	2000					
CGTTTGACGT GCAAACTGCA	TATGTGTGAA ATACACACTT	AGCGTCAATA TCGCAGTTAT	ACTAAACCTT TGATTTTGAA	GGATTGACTA CCTPACTGAT	GTCAAAATAT CAGTTTATA	AAACTGATTG TTTGACTAAC	CATTGAATTC GTAACCTTAAG	TTGAAAATTT AACTTTTAAA	TCCCTTAAAA AGGGAATTTT
2020	2040	2060	2080	2100					
TGAACATGAA ACTTGTACTT	TTTCATCAAG AAAGTAGTTC	ATTTTGTCTT TAAACAGAAA	TTGGAAGGAT AACCTTCTCA	GTGATTTATA CACTAAATAT	ATCTATACAA TAGATATGTT	TCATACATTT AGTATGTAAA	TGCTATGATAT ACGTACTATA	TAGTTTTTTG ATCAAAAAAC	AAGAAGCTAA TTCTTGGTTT
2120	2140	2160	2180	2200					
AATAGAGCTT TTATCTCGAA	CTTTATAAAA GAAATATTTT	CTGATTAGCG GACTAAATCG	CTTGATAAGA GAACTATTCT	AAAAGAAGGT TTTCTTCCCA	AGATAATCGA TCTATTAGCT	ACTCATGGGG TGAGTACCCC	ATGAGTTAAA TACTCAATTT	AATGTGTGCA TTACACACGT	CTTAGTTTCT GAATCAAAGA
2220	2240	2260	2280	2300					
AAAACCTTTT TTTTGGAAAA	GAAGTCGAAA CTTCAGCTTT	CAATGACAAT GTCTACTGTTA	ATTGCTGCGC TAACGCAGCG	AAGTTGATAT TTCAACTATA	ATAACAGGAT TATTGTCTCA	CTTAAAGTTG GAAATTTCAAC	AAATGTGAAA TTTAACTATT	TTGAGTTTCT AAGTCTAAAA	AAATTTTAGAG TTAAAAATCTC
2320	2340	2360	2380	2400					
CACCAGATGA GTGGTCTACT	TCAGAGTTTC AGTCTCAAGG	AGATTTACAT TCTAAATGTA	TTGAAGTATA AACTTCATAT	AAACATTTTG TTTGTAAAAC	AACACATATA TTGTGTATAT	TCTAAAGCAG AGATTTCGTC	TAACTTCAAA ATTGAAGTTT	AATAGGGTAA TTATCCCAAT	CTAATAGTAA GATTATCAAT
2420	2440	2460	2480	2500					
CTTACATTGT GAATGTAACA	TTTTTTTAA AAAAAATTTA	GCTTTTATAC CGAAAAATAT	TTACTATCAT AATGATAGTA	TTTTATATAT AAAAATATATA	AGATGCCTGG TCTACGGACC	TTAAGTAAAG AATTCATTTC	ATGATTATCA TACTAATAGT	AAAACCTGTT TTTTGACAAC	GTTAGTAAAC CAATCATTTGT
2520	2540	2560	2580	2600					
GAAATTTGTT CTTTAAACAC	CAAAATGTAAC GTTTACATTG	ATATTATATA TATAATATAT	AGCTTTCTTT TCGAAAGAAA	CACCTTGGTG GTGAAACCCAC	CATTCTCTCT GTAAGAGAGA	AAATAATGGC TTTATTACCG	CTCTATTGAT GAGATAACTA	GCAGTATCTG CGTCATAGAC	ATTCCTTAGTT TAAGAATCAA
2620	2640	2660	2680	2700					
TTGAAATGGT AACTTTACCA	TTTTGCATAA AAAACGTATT	ATTATTGTTT TAATACAAGG	TAATGATATT ATTACGTAAA	TTGTTTTATC AACAATAATAG	TCCAGCATTT AGGTCGTAAA	CCAAGATCAT GGTCTAGTAA	TGATCGTTAT ACTAGCAATA	GAAATACAA CTTTATGTTG	ATGCTGATCA TACGACTACT
2720	2740	2760	2780	2800					
ACTTAGAGCC TGAATCTCGG	TTAGTAAAGT AATCATTTAT	ATTAGCTAAG TAATCGATTG	AACGTCATT TTGCAGTAAG	TAATATTCTT ATTATAAGAA	CTGGATCGCG GACCTACGCC	TTTTTGGTGT AAAAACCACA	TATGAAGGAT ATACTTCCTA	AGAAGCGCTG TCTTCGGCAG	TTCAAGCCCG AAGTTGCGCC
2820	2840	2860	2880	2900					
AGAAACCTCA TCTTTGGAGT	ATGTTTGGAA TACAAAACCT	CTCGTAACAC GAGCATTTGT	CGAACTTAAT GCTTGAAATTA	TCTCTAGAGT AGAGATCTCA	TACAGTTATT ATGTCAATAA	GTGCTCTACTG CACAGATGAC	GAAATACAA CTTTTATGTT	GAACTTCACA CTTGAAAGTT	ATCTTTCTGA TAGAAGAGCT
2920	2940	2960	2980	3000					
CCATTCCCTT GGTAAGGAAA	TCTTCAATGT AGAAGTACAC	CAGGATCTTG GTCCTAGAAC	AGAAAGAAAT TTCTTTTTTA	TCGAAATTTAT AGTCTTAAATA	CTTCCACACA GAAGGTGTGT	AGGAGTTACT TCCTCAATGA	AGAAACCTAG TCTTTGTTCAG	CAAAGGTTAG GTTTCCAATC	CAGTACGACA GTCATGCTGT
3020	3040	3060	3080	3100					
CATTTTTCTC GTAAAAAGAG	CCCTCTTCTT GGGAGAGAAG	CTGATAAAAA GACTATTTTT	AAATGTTTTT TTTACAAAAA	TTTCTTTTGT AAGAAAAACA	CTACTTGTGA GATGAACACT	ATACAGCAAG TATGTCTGTC	CTTGAAGGAT GAACTTCTTG	CAAATGTGCA GTTTACAGCT	TAATGTAAAGT ATTACATTCA
3120	3140	3160	3180	3200					
GTAGATTCTC CATCTAAGAG	TAATTTCTCT ATTAAAGAGA	GGAGGAACAA CCTCCTTGTG	CTTGAGACTG GAACTCTGAC	CTCTGTCCGT GAGACAGGCA	AAGTAGAGCT TTCATCTCGA	AGGAAGGTAT TCCTTCCATA	ATGTGCTGCT TACACGACGA	ACTAAGTGAT TGATTCACTA	TCAACCAATT AGTTGGTTAA
3220	3240	3260	3280	3300					
ACTCCACAAA TGAGGTGTTT	ACCTTCTTTT TGGAAAGAAA	TAGTTAGTTA ATCAATCAAT	TCCTAGAACA AGGATCTTGT	ATCTTTTGAC TAGAAAACTG	ATAAATCTTA TATTTAGAAT	ATGTCTTGT TACAGAACAA	ATAGGCAGAA TATCCGCTCT	CTGATGATGG GACTACTACC	AGTATATCGA TCATATAGCT
3320	3340	3360	3380	3400					
GTCCCTTAAA CAGGGAATTT	GAAAAGGTTA CTTTTCCAAAT	GTGCTTTGGT CACGAAACCA	TTTATTTTTC AAAATAAAAG	GATAAAGGCC CTATTTCGGG	ATAITCTAGG TATAAGATCC	CTATGATGAT GATACTACTA	TCTTGAATTC AGAACCTTAAG	TATTAACCTG ATAATTGAGC	CTGAGTCTAC GACTCAGATG
3420	3440	3460	3480	3500					
AGATTACTAT TCTAATGATA	ATATATATAT TATATAGAAA	ATATATCTTT TATATAGAAA	TGGTCTTGTG ACCAGAACAG	TTAGTTCTCT AATCAAGGAC	ATTAGTATT TAAATCATAA	GGCTTCATTC CCGAAGTAAG	AGGTGAACCC TCCACTTTGG	CTAATGAGAA GATTACTCTT	TTAAAAAAC AATTTTTTTG
3520	3540	3560	3580	3600					
AAGCAGTTTT TTCGTCAAAA	AAACTCTTGA TTTGAGAACT	TCAAATCCAA AGTTTAGGTT	CCTTTCCCTC GGAAAGGGAG	ATAAAGTGTC TATTTCAACG	GAAATTTGAT CTTAAACCTA	GAGGATGATT CTCCTACTAA	TATGTTTCGA ATACAAAGCT	GAAGGAAACA CTTCCTTTGT	TGTTTTGGAAA ACAAACCTTT
3620	3640	3660	3680	3700					
TAGCTATAGA ATCGATATCT	AGTTGTTTGA TCAACAATCT	AACTAATGAC TTGATTACTG	CTTATGATCT GAATACTAGA	TTTCCAAACA AAAGGTTTGT	GGAGAAATTT CCTCTTTAAC	CTGAGAGAAG GACTCTCTTC	AGAACCAGGT TCTTGGTCCA	TCTGGCTAGC AGACCGATCG	CAGGTAACAA GTCCATTGTT
3720	3740	3760	3780	3800					
TGACCACAAT ACTGGTGTTA	ATCTTCTGCT TAGAAGACGA	CTTGAAGCTA GAACTTCGAT	ATTAATCACT TAATTAGTGA	TTATACGTCC AATATGACAG	CCGTTATAGA GGCAATATCT	GAGATACACA CTCTATGTGT	TATACACGTA ATATGTGCAT	CATGAAAAC GTACTTTTGA	AAAAGTTGAA TTTTCACATT
3820	3840	3860	3880	3900					
GGACTTTGAT CCTGAAACTA	GGATACCTAGA CCTATGATCT	CAATTATAGT GTTAATATCA	GAAACCTCAA CTTTGGGATT	ATATGTGATA TATACACTAT	AGTGATAACA TCACTATTTG	AAATGCTTTT TTTACGAAAA	AAAATCTATC TTTTAGATAG	TTTCTTGTTA AAAGAACAAAT	ATTTAGTAGC TAAATCATCG

```

3920      3940      3960      3980      4000
TGTCAGAGAA GAAAGGTATG TCTCACCAGT GAAAGATACT CAAAACCCGG TATTTTAAAT TTGTGAAATT TGCAAAATAA AAAAATGCTT TCTACAAGAT
ACAGTCTCTT CTTTCCATAC AGAGTGGCTA CTCTTCTATGA GTTTTGGGCC ATAAAAATTA AACACITTTAA ACGTTTATTT TTTTACGAA AGATGTTCTA

4020      4040      4060      4080      4100
AGATTAATTT CTTGCAATGT TTAGTAGCTG TAGAAAAAAA AGAAATGTAA GAAAGTTTCT TACAGATGGG AAAGAATACG TTGCTGGCAA CAGATGATGA
TCTAATTAAA GAACGTTTACA AATCATCGAC ATCTTTTITT TCTTTACATT CTTTCAAAGA ATGCTTACCC TTTCTTATGC AACGACCGTT GTCTACTACT

4120      4140      4160      4180      4200
GAGAGGAATG TTTCCGGGAA GTAGTCCCGG CAACAAAATA CCGGAGACTC TCCCGCTGCT CAATTAGCCA CCATCATCAA CGGCTGAGTT TTCACCTTAA
CTCTCCTTAC AAAGGCCCTT CATCGAGGCC GTTGTTTTAT GGCCTCTGAG AGGGCGACGA GTTAATCGGT GGTAGTAGTT GCCGACTCAA AAGTGAATTT

```

SEQ ID NO:48 and SEQ ID NO:49

Alternatively splice Arabidopsis AGL27 cDNA and resulting Alternate Arabidopsis AGL27 amino acid sequence

```

20      40      60      80      100
ATGSGAAGAA GAAAAATCGA GATCAAGCGA ATCGAGAACA AAAGCAGTCG ACAAGTCACT TTCTCCAAAC GACGCAATGG TCTCATCGAC AAAGCTCGAC
TACCCCTTCTT CTTTTCGAGT CTAGTTCGCT TAGCTCTTGT TTTCTGCAGC TGTTTCAGTGA AAGAGGTTTG CTGCGTTACC AGAGTAGCTG TTTGAGCTG
M G R R K I E I K R I E N K S S R Q V T F S K R R N G L I D K A R>

120      140      160      180      200
AATTTTCGAT TCTCTGTGAA TCCTCCGTCG CTGTTGTCTG CGTATCTGCC TCCGGAACAC TCTATGACTC TTCTCCGGT GACGAGATAG AAGCGCTGTT
TTGAAAGCTA AGAGACACTT AGGAGGCGAG GACAACAGCA GCATAGACGG AGGCCTTTTG AGATACTGAG AAGGAGGCCA CTGCTCTATC TTCGCGACAA
Q L S I L C E S S V A V V V V S A S G K L Y D S S S G D E I E A L F>

220      240      260      280      300
CAAGCCGGAG AAACCTCAAT GTTTTGAAGT CGATCTTGAA GAAAAAATTC AGAATTATCT TCCACACAAG GAGTTACTAG AAACAGTCCA AAGCAAGCTT
GTTCCGGCTC TTTGGAGTTA CAAAACITGA GCTAGAAGCTT CTTTTTTAAG TCTTAATAGA AGGTGTGTTT CTCAATGATC TTTGTTCAGT TTCGTTGCAA
K P E K P Q C F E L D L E E K I Q N Y L P H K E L L E T V Q S K L>

320      340      360      380      400
GAAGAACCAA ATGTCGATAA TGTAAGTGTA GATTCTCTAA TTTCTCTGGA GGAACAACTT GAGACTGCTC TGTCCGTAAG TAGAGCTAGG AAGGCAGAAC
CTTCTTGTTT TACAGCTATT ACATTACATC CTAAGAGATT AAAGAGACCT CCTTGTGTA CTTCTGACGAG ACAGGCATTC ATCTCGATCC TTCGCTCTTG
E E P N V D N V S V D S L I S L E E Q L E T A L S V S R A R K A E>

420      440      460      480      500
TGATGATGGA GTATATCGAG TCCCTTAAAG AAAAGGAGAA ATTGCTGAGA GAAGAGAACC AGTTTCTGGC TAGCCAGATG GGAAAGAATA CGTTGCTGGC
ACTACTACCT CATATAGCTC AGGGAATTTT TTTTCTCTTT TAACGACTCT CTTCTCTTGG TCCAAGACCG ATCGGTCTAC CCTTTCTTAT GCAACGACCG
L M M E Y I E S L K E K E K L L R E E N Q V L A S Q M G K N T L L A>

520      540      560
AACAGATGAT GAGAGAGGAA TGTTCCTGGG AAGTAGCTCC GGCAACAAAA TACCGGAGAC TCTCCCGCTG CTCAATTAG
TTGTCTACTA CTCTCTCCTT ACAAAGGCC TTTATCGAGG CCGTTGTTTT ATGGCCTCTG AGAGGGCGAC GAGTTAATC
T D D E R G M F P G S S S G N K I P E T L P L L N *>

```